

Query Match 3.6%; Score 7; DB 2; Length 729;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 IKKEQDK 177

Db 359 IKKEQDK 365

RESULT 37

I59282

diacylglycerol kinase (EC 2.7.1.107) gamma - rat

C:Species: Rattus sp. (rat)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: I59282

R:Goto, K.; Funayama, M.; Kondo, H.

A:Title: Cloning and expression of a cytoskeleton-associated diacylglycerol kinase that

A:Reference number: I59282; MUID:95108095; PMID:7809169

A:Accession: I59282

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-788 <RES>

A:Cross-references: GB:D38448; NID:q643598; PIDN:BAA07480.1; PID:q784935

C:Superfamily: human diacylglycerol kinase; calmodulin repeat homology; protein kinase C

C:Keywords: ATP; calcium binding; duplication; EF hand; phosphotransferase; zinc

F:172-204/Domain: calmodulin repeat homology <EF1>

F:217-249/Domain: calmodulin repeat homology <EF2>

F:269-318/Domain: protein kinase C zinc-binding repeat homology <KZ1>

F:334-380/Domain: protein kinase C zinc-binding repeat homology <KZ2>

Query Match

Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRK 134

Db 781 RRKRSRK 787

RESULT 38

A53691

diacylglycerol kinase (EC 2.7.1.107) gamma - human

C:Species: Homo sapiens (man)

C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 20-Jun-2000

C:Accession: A53691

R:Kai, M.; Sakane, F.; Imai, S.; Wada, I.; Kanoh, H.

J. Biol. Chem. 269, 18492-18498, 1994

A:Title: Molecular cloning of a diacylglycerol kinase isozyme predominantly expressed in

A:Reference number: A53691; MUID:94308084; PMID:8034597

A:Accession: A53691

A:Molecule type: mRNA

A:Residues: 1-791 <KAI>

A:Cross-references: GB:D26135; NID:q516757; PIDN:BAA05132.1; PID:q516758

C:Genetics:

A:Gene: GDB:DAGK3

A:Cross-references: GDB:377343

A:Map position: 12q13.3-12q13.3

C:Superfamily: human diacylglycerol kinase; calmodulin repeat homology; protein kinase C

C:Keywords: ATP; calcium binding; duplication; EF hand; phosphotransferase; zinc

F:175-207/Domain: calmodulin repeat homology <EF1>

F:220-252/Domain: calmodulin repeat homology <EF2>

F:272-321/Domain: protein kinase C zinc-binding repeat homology <KZ1>

F:337-383/Domain: protein kinase C zinc-binding repeat homology <KZ2>

Query Match

Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRK 134

Db 784 RRKRSRK 790

RESULT 39

JN0878

100K protein - fowl adenovirus 10

C:Species: Aviadenovirus gallio (fowl adenovirus 10)

C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 08-Oct-1999

C:Accession: JN0878

R:Sheppard, M.

Gene 132, 307-308, 1993

A:Title: Identification of a fowl adenovirus gene with sequence homology to the 100K

A:Reference number: JN0878; MUID:94040780; PMID:8224879

A:Accession: JN0878

A:Molecule type: DNA

A:Residues: 1-798 <SHE>

A:Cross-references: GB:L07890; NID:q348034; PIDN:AAA72328.1; PID:q348035

A:Experimental source: serotype 10

C:Genetics:

A:Gene: 100K

A:Map position: 57.3 to 62.6 map units

C:Superfamily: adenovirus late 100K protein

Query Match

Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ENPNPRD 117

Db 592 ENPNPRD 598

RESULT 40

C84615

hypothetical protein At2g22660 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: C84615

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: C84615

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-819 <STO>

A:Cross-references: GB:AE002093; NID:q4314363; PIDN:AAD15574.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g22660

A:Map position: 2

Query Match

Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AKOHLFF 188

Db 27 AKOHLFF 33

Search completed: February 26, 2003, 10:01:52

Job time : 71 secs

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 VTPEMLN 11  
|||||||  
Db 68 VTPEMLN 74

## RESULT 32

S64250  
Probable membrane protein YGL228w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein G0954  
C:Species: Saccharomyces cerevisiae  
C>Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 19-Apr-2002  
C:Accession: S64250  
R:Rattmann, B.; Kramer, B.; Kramer, W.  
Submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64248  
A:Accession: S64250  
A:Molecule type: DNA  
A:Residues: 1-577 <FAR>  
A:Cross-references: EMBL:Z72750; NID:gl322882; PID:e243889; PID:gl322883; GSPDB:GN000007;  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:SHE10; MIPS:YGL228w  
A:Cross-references: SGD:S0003197  
A:Map position: 7L  
C:Keywords: transmembrane protein  
F;2-18/Domain: transmembrane tstatus predicted <TM>

Query Match 3.6%; Score 7; DB 2; Length 577;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 KYDFIVG 65  
|||||||  
Db 175 KYDFIVG 181

## RESULT 33

T16774  
Hypothetical protein SSSD1.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16774  
R:Minx, P.  
Submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid SSSD1.  
A:Reference number: Z18575  
A:Accession: T16774  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-599 <MIN>  
A:Cross-references: EMBL:U41530; NID:gl118030; PID:gl118031; PIDN:AAA83273.1; CESP:SSSD1  
C:Genetics:  
A:Gene: CESP:SSSD1.1  
A:Introns: 28/1; 67/3; 107/3; 125/3; 142/2; 180/3; 230/2; 262/1; 302/3; 343/3; 370/1; 38

Query Match 3.6%; Score 7; DB 2; Length 599;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 KRRKRS 133  
|||||||  
Db 575 KRRKRS 581

## RESULT 34

B86789  
Hypothetical protein pabB [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C:Species: Lactococcus lactis subsp. lactis  
C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: B86789  
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: B86789  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-628 <STO>  
A:Cross-references: GB:AE005176; PID:gl2724293; PIDN:AAK05412.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: pabB

Query Match 3.6%; Score 7; DB 2; Length 628;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 FSEILK 59  
|||||||  
Db 413 FSEILK 419

## RESULT 35

F84996  
Phosphotransferase system enzyme II (EC 2.7.1.69) [imported] - Buchnera sp. (strain A  
N:Alternate names: pts system mannitol-specific IIABC component  
C:Species: Buchnera sp.  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: F84996  
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.  
Nature 407, 81-86, 2000  
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp  
A:Reference number: A84930; MUID:20445173; PMID:10993077  
A:Accession: F84996  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-632 <STO>  
A:Cross-references: GB:AP000398; GSPDB:GN00144  
A:Experimental source: strain APS  
C:Genetics:  
A:Gene: mtIIA; B0572  
C:Superfamily: phosphotransferase system mannitol-specific enzyme II; phosphotransfer  
actor III homology  
C:Keywords: phosphotransferase

Query Match 3.6%; Score 7; DB 2; Length 632;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 ILLKYDF 62  
|||||||  
Db 332 ILLKYDF 338

## RESULT 36

A97747  
Oligopeptidase B (EC 3.4.21.83) [imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 17-May-2002  
C:Accession: A97747  
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;  
Science 293, 2093-2098, 2001  
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: A97747  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-729 <KUR>  
A:Cross-references: GB:AE006914; PIDN:AAL02915.1; PID:gl15619442; GSPDB:GN00173  
C:Genetics:  
A:Gene: ptrB  
C:Superfamily: prolyl oligopeptidase  
C:Keywords: hydrolase; serine proteinase



Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 EKSAPDV 46  
|||||||

Db 347 EKSAPDV 353

RESULT 28

T20382

hypothetical protein D2089.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T20382; Z19264

R:Swinsburne, J.

submitted to the EMBL Data Library, September 1994

A:Reference number: Z19264

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Accession: T20382

A:Molecule type: DNA

A:Residues: 1-480 <WIL>

A:Cross-references: EMBL:Z36948; PIDN:CAA85414.1; GSPDB:GN00020; CESP:D2089.1

A:Experimental source: clone D2089

R:Gardner, A.

submitted to the EMBL Data Library, June 1995

A:Reference number: Z19803

A:Accession: T23815

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-480 <W12>

A:Cross-references: EMBL:Z49911; PIDN:CAA90135.1; GSPDB:GN00020; CESP:D2089.1

A:Experimental source: clone M28

C:Genetics:

A:Gene: CESP:D2089.1

A:Map position: 2

A:Introns: 40/3; 91/1; 132/1; 158/3; 217/3; 262/2

Query Match 3.6%; Score 7; DB 2; Length 480;

Best Local Similarity 100.0%; Pred. No. 31;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 IKFDDER 124  
|||||||

Db 59 IKFDDER 65

RESULT 29

S67625

probable membrane protein YDL089w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D2416

C:Species: Saccharomyces cerevisiae

C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002

C:Accession: S67625; S67631; S67428

R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67608

A:Accession: S67625

A:Molecule type: DNA

A:Residues: 1-484 <WAM>

A:Cross-references: EMBL:Z74137; NID:g1431117; PID:g1431118; MIPS:YDL089w

A:Experimental source: strain S288C

R:Ballesta, J.P.G.; Remacha, M.; Soler-Mira, A.; Jimenez, A.; Garcia-Cantalejo, J.M.; Bo

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67629

A:Accession: S67631

A:Molecule type: DNA

A:Residues: 1-484 <BAL>

A:Cross-references: EMBL:Z74137; NID:g1431117; PID:g1431118; MIPS:YDL089w

A:Experimental source: strain S288C

R:Boskovic, J.; Saiz, J.E.; Soler-Mira, A.; Jimenez, A.; Garcia-Cantalejo, J.L.; Jimenez

submitted to the EMBL Data Library, February 1996

A:Reference number: S67406

A:Accession: S67428

A:Molecule type: DNA

A:Residues: 1-484 <BOS>

A:Cross-references: EMBL:X95644; NID:g1199535; PID:g1199558

C:Genetics:

A:Cross-references: SGD:S0002247

A:Map position: 4L

A>Note: YDL089w

C:Superfamily: Saccharomyces cerevisiae probable membrane protein YDL089w

C:Keywords: transmembrane protein

F:68-84/Domain: transmembrane #status predicted <TM1>

F:147-163/Domain: transmembrane #status predicted <TM2>

F:241-257/Domain: transmembrane #status predicted <TM3>

Query Match 3.6%; Score 7; DB 2; Length 484;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 QLVINEK 41  
|||||||

Db 42 QLVINEK 48

RESULT 30

T29814

hypothetical protein C46A5.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T29814

R:Johnson, D.; Stellyes, L.

submitted to the EMBL Data Library, June 1996

A:Description: The sequence of C. elegans cosmid C46A5.

A:Reference number: Z20690

A:Accession: T29814

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-497 <JOH>

A:Cross-references: EMBL:U61948; PIDN:AA03142.1; GSPDB:GN00022; CESP:C46A5.1

A:Experimental source: strain Bristol N2; clone C46A5

C:Genetics:

A:Gene: CESP:C46A5.1

A:Map position: 4

A:Introns: 85/3; 108/1; 189/1; 221/3; 329/2; 367/2; 444/3; 473/3

Query Match 3.6%; Score 7; DB 2; Length 497;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRSKS 135  
|||||||

Db 35 RKSRSKS 41

RESULT 31

A87627

PDZ domain family protein [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: A87627

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

n, J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: A87627

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-561 <STO>

A:Cross-references: GB:AE005673; NID:g13424697; PIDN:AAK25013.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3051

Query Match 3.6%; Score 7; DB 2; Length 561;

Best Local Similarity 100.0%; Pred. No. 36;

C>Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 20-Sep-1999  
C:Accession: S03762  
R:Britton, P.; Carmenes, R.S.; Page, K.W.; Garwes, D.J.; Parra, F.  
Mol. Microbiol. 2, 89-99, 1988  
A:Title: Sequence of the nucleoprotein gene from a virulent British field isolate of tra  
A:Reference number: S03762; MUID:86216185; PMID:2835592  
C:Accession: S03762  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-382 <BRI>  
A:Cross-references: GB:Y00542; NID:g59012; PIDN:CAA68607.1; PID:g59013  
C:Superfamily: coronavirus nucleocapsid protein  
C:Keywords: nucleocapsid

Query Match 3.6%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135  
|||||

Db 342 RKSRKS 348

RESULT 24

E36607  
Nucleocapsid protein - porcine respiratory virus  
N:Alternate names: N protein  
C:Species: porcine respiratory virus  
C>Date: 30-Aug-1991 #sequence\_revision 30-Aug-1991 #text\_change 20-Sep-1999  
C:Accession: E36607; S10447  
R:Rasschaert, D.; Duarte, M.; Laude, H.  
J. Gen. Virol. 71, 2599-2607, 1990  
A:Title: Porcine respiratory coronavirus differs from transmissible gastroenteritis virus  
A:Reference number: A36607; MUID:91073120; PMID:2174956  
C:Accession: E36607  
A:Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-382 <RAS>  
A:Cross-references: GB:224675; NID:g395057; PIDN:CAA80841.1; PID:g395062  
A:Experimental source: strain RM4  
R:Page, K.W.; Britton, P.; Boursnell, M.E.G.  
submitted to the EMBL Data Library, March 1990  
A:Reference number: S10447  
A:Accession: S10447  
A:Molecule type: mRNA  
A:Residues: 1-23 <PAC>  
A:Cross-references: EMBL:X52668; NID:g58981; PIDN:CAA36895.1; PID:g58982  
C:Superfamily: coronavirus nucleocapsid protein  
C:Keywords: nucleocapsid

Query Match 3.6%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135  
|||||

Db 342 RKSRKS 348

RESULT 25

S47428  
Nucleocapsid protein - porcine transmissible gastroenteritis virus  
N:Alternate names: N protein  
C:Species: porcine transmissible gastroenteritis virus  
C>Date: 23-Nov-1994 #sequence\_revision 19-Apr-1996 #text\_change 20-Sep-1999  
C:Accession: S65856; S47428  
R:Chen, C.M.; Cavanagh, D.; Britton, P.  
Virus Res. 38, 83-89, 1995  
A:Title: Cloning and sequencing of a 8.4-kb region from the 3'-end of a Taiwanese virul  
A:Reference number: S65850; MUID:96060227; PMID:8546012  
A:Accession: S65856  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: genomic RNA

A:Residues: 1-382 <CH2>  
A:Cross-references: EMBL:Z35758; NID:g529246; PIDN:CAA84811.1; PID:g529253  
A:Experimental source: Taiwanese field isolate  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
C:Genetics:  
A:Gene: N  
C:Superfamily: coronavirus nucleocapsid protein  
C:Keywords: nucleocapsid

Query Match 3.6%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135  
|||||

Db 342 RKSRKS 348

RESULT 26

F75596  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: F75596  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: F75596  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-405 <WHI>  
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12358.1; PID:g646  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0036  
A:Map position: 2

Query Match 3.6%; Score 7; DB 2; Length 405;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 SEILKY 60  
|||||

Db 234 SEILKY 240

RESULT 27

B70185  
Probable hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C>Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 05-May-2000  
C:Accession: B70185  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: B70185  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-407 <LE>  
A:Cross-references: GB:AE001169; GB:AE000783; NID:g2688611; PIDN:AAC67034.1; PID:g268  
A:Experimental source: strain B31  
C:Superfamily: hydroxymethylglutaryl-CoA synthase  
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 3.6%; Score 7; DB 2; Length 407;  
Best Local Similarity 100.0%; Pred. No. 27;

F:111-173/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 3.6%; Score 7; DB 2; Length 344;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135  
|||||

Db 241 RKSRKS 247

#### RESULT 19

AC3224

hypothetical protein Atu523 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: AC3224

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AC3224

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 <KUR>

A:Cross-references: GB:AE008687; PIDN:AAL46209.1; PID:g17743983; GSPDB:GN00188

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu523

A:Genome: plasmid

#### Query Match

Best Local Similarity 100.0%; Pred. No. 25;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 LRLRGFY 78

|||||

Db 117 LRLRGFY 123

#### RESULT 20

I39939

phenylalanine dehydrogenase (EC 1.4.1.-) - Bacillus sphaericus

C:Species: Bacillus sphaericus

C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 11-Jun-1999

C:Accession: I39939

R:Okazaki, N.; Hibino, Y.; Asano, Y.; Ohmori, M.; Numao, N.; Kondo, K.

Gene 63, 337-341, 1988

A:Title: Cloning and nucleotide sequencing of phenylalanine dehydrogenase gene of Bacill

A:Reference number: I39939; MUID:88255879; PMID:2838396

A:Accession: I39939

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-381 <RES>

A:Cross-references: GB:M26661; NID:g341510; PIDN:AAA22646.1; PID:g529017

C:Genetics:

A:Gene: pdh

C:Superfamily: leucine dehydrogenase

C:Keywords: oxidoreductase

#### Query Match

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 IVKSDDI 32

|||||

Db 236 IVKSDDI 242

#### RESULT 23

S03762

nucleocapsid protein - porcine transmissible gastroenteritis virus (strain FS772/70)

C:Species: porcine transmissible gastroenteritis virus

#### RESULT 21

VH1HPC

nucleocapsid protein - porcine transmissible gastroenteritis virus

C:Species: porcine transmissible gastroenteritis virus

C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 29-Oct-1999

C:Accession: A04025; S03936

R:Kapke, P.A.; Brian, D.A.

Virolgy 151, 41-49, 1986

A:Title: Sequence analysis of the porcine transmissible gastroenteritis coronavirus n

A:Reference number: A94340; MUID:86181608; PMID:3008432

A:Accession: A04025

A:Molecule type: genomic RNA

A:Residues: 1-382 <KAP>

A:Cross-references: GB:M14878; NID:g335209; PIDN:AAA47915.1; PID:g335211

A:Experimental source: strain Purdue

R:Rasschaert, D.; Gelfi, J.; Laude, H.

Biochimie 69, 591-600, 1987

A:Title: Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organizat

A:Reference number: S01738; MUID:88078100; PMID:2825819

A:Accession: S03936

A:Molecule type: mRNA

A:Residues: 1-236, 'w', 238-375, 'N', 377-382 <RAS>

A:Cross-references: EMBL:X06371; NID:g58995; PIDN:CAA29674.1; PID:g59001

A:Experimental source: strain Purdue-115

C:Superfamily: coronavirus nucleocapsid protein

C:Keywords: glycoprotein; nucleocapsid

F:134,154,172,267,364/Binding site: carbohydrate (Asn) (covalent) #status predicted

#### Query Match

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135

|||||

Db 342 RKSRKS 348

#### RESULT 22

S24282

nucleocapsid protein - porcine respiratory virus (strain 86/137004)

N:Alternate names: N protein

C:Species: porcine respiratory virus

A:Variety: strain 86/137004

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999

C:Accession: S24282; S21311

R:Britton, P.; Mawditt, K.L.; Page, K.W.

Virus Res. 21, 181-198, 1991

A:Title: The cloning and sequencing of the virion protein genes from a British isolat

A:Reference number: S24279; MUID:92116634; PMID:1662846

A:Accession: S24282

A:Molecule type: genomic RNA

A:Residues: 1-382 <BR1>

A:Cross-references: EMBL:X60056; NID:g61342; PIDN:CAA42657.1; PID:g61346; EMBL:X55980

A:Experimental source: strain 86/137004

C:Genetics:

A:Gene: N

C:Superfamily: coronavirus nucleocapsid protein

C:Keywords: glycoprotein; nucleocapsid

F:134,154,172,267,364/Binding site: carbohydrate (Asn) (covalent) #status predicted

#### Query Match

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135

|||||

Db 342 RKSRKS 348

C:Genetics:  
A:Gene: eq\_1343

Query Match 3.6%; Score 7; DB 2; Length 160;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 RLEDYIK 98  
|||||||  
Db 151 RLEDYIK 157

## RESULT 14

E71407

hypothetical protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

A:Variety: columbia

C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998

C:Accession: E71407

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks

P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel

avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.

Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech

erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans

C.; Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal

A:Reference number: A71400; MUID:98121113; PMID:9461215

C:Accession: E71407

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-194 &lt;BEV&gt;

C:Cross-references: GB:257336; NID:g2244788; PID:e326896; PID:g2244808

C:Genetics:

A:Map position: 4COP9-4G3845

## Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 194;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 ERPHRR 129

|||||||

Db 160 ERPHRR 166

## RESULT 15

T05238

hypothetical protein F18A5.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999

C:Accession: T05238

R:Bevan, M.; Weber, N.; Grueninger, D.; Schmidheini, T.; Bancroft, I.; Mewes, H.W.; Maye

submitted to the Protein Sequence Database, February 1999

A:Reference number: Z15405

C:Accession: T05238

A:Molecule type: DNA

A:Residues: 1-194 &lt;BEV&gt;

A:Cross-references: EMBL:AL035528

A:Experimental source: cultivar Columbia; BAC clone F18A5

C:Genetics:

A:Map position: 4

A:Introns: 13/2; 108/1; 154/1

A:Note: F18A5.80

## Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 194;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRKD 162

|||||||

Db 98 TSKRRKD 104

## RESULT 16

T47871

hypothetical protein T4C21.10 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T47871

R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;

W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, March 2000

A:Reference number: 224479

A:Accession: T47871

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-250 &lt;CHO&gt;

A:Cross-references: EMBL:AL162295

A:Experimental source: cultivar Columbia; BAC clone T4C21

C:Genetics:

A:Map position: 3

A:Introns: 34/1; 57/3; 85/1; 123/3; 171/3; 187/3; 213/3

A:Note: T4C21.10

## Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 250;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 KEVTPEM 9

|||||||

Db 117 KEVTPEM 123

## RESULT 17

JE0175

freezed protein-lb - human

C:Species: Homo sapiens (man)

C:Date: 03-Jul-1998 #sequence\_revision 10-Jul-1998 #text\_change 07-May-1999

C:Accession: JE0175

R:Hu, E.; Zhu, Y.; Fredrickson, T.; Barnes, M.; Kelsell, D.; Beeley, L.; Brooks, D.

Biochem. Biophys. Res. Commun. 247, 287-293, 1998

A:Title: Tissue restricted expression of two human frzbs in preadipocytes and pancrea

A:Reference number: JE0174; MUID:98308108; PMID:9642118

C:Accession: JE0175

A:Molecule type: mRNA

A:Residues: 1-317 &lt;HUA&gt;

C:Genetics:

A:Gene: hFRP-1b

A:Map position: 5ql4.3-ql2.1

## Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 317;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 159 KRKDKR 165

|||||||

Db 236 KRKDKR 242

## RESULT 18

S59043

spilling factor SRP55 - human

C:Species: Homo sapiens (man)

C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 28-May-1999

C:Accession: S59043

R:Scraton, G.R.; Caceres, J.F.; Mayeda, A.; Bell, M.V.; Plebanski, M.; Jackson, D.G.

EMBO J. 14, 4336-4349, 1995

A:Title: Identification and characterization of three members of the human SR family

A:Reference number: S59042; MUID:96016206; PMID:7556075

C:Accession: S59043

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-344 &lt;SCR&gt;

A:Cross-references: GB:U30883; NID:g1049087; PIDN:AAA93073.1; PID:g1049088

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleopro

F:3-62/Domain: ribonucleoprotein repeat homology &lt;RRM1&gt;

A:Accession: S64928  
A:Molecule type: DNA  
A:Residues: 1-502 <BEN>  
A:Cross-references: EMBL:273266; NID:g1360470; PIDN:CAA97656.1; PID:g1360471; GSPDB:GN0004  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:GIS3; MIPS:YLR094c  
A:Cross-references: SGD:S0004084  
A:Map position: 12R  
C:Superfamily: Saccharomycetes cerevisiae hypothetical protein YLR094c

Query Match 4.2%; Score 8; DB 2; Length 502;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRKS 135  
|||||  
Db 281 RRKRSRKS 288

RESULT 10  
A57459  
ribosomal protein S6 kinase 2 (EC 2.7.1.-) 2 - human  
N:Alternate names: growth factor-stimulated protein kinase (GFSPK); p90 S6 kinase; ribos  
C:Species: Homo sapiens (man)  
C:Date: 16-Feb-1996 #sequence\_revision 16-Feb-1996 #text\_change 18-Jun-1999  
C:Accession: A57459; I64834  
R:Zhao, Y.; Bjorbaek, C.; Weremowicz, S.; Morton, C.C.; Moller, D.E.  
Mol. Cell. Biol. 15, 4353-4363, 1995  
A:Title: RSK3 encodes a novel pp90rsk isoform with a unique N-terminal sequence: growth  
A:Reference number: A57459; MUID:95349602; PMID:7623830  
A:Accession: A57459  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-733 <ZHA>  
A:Cross-references: EMBL:X85106; NID:g1033032; PIDN:CAA59427.1; PID:g1033033  
R:Moller, D.E.; Xia, C.H.; Tang, W.; Zhu, A.X.; Jakubowski, M.  
Am J. Physiol. 266, 351-359, 1994  
A:Title: Human rsk isoforms: cloning and characterization of tissue-specific expression.  
A:Reference number: I51901  
A:Accession: I64834  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 'A', 1-255, 'A', 257-268, 'S', 270-446, 'G', 448-539 <MOL>  
A:Cross-references: GB:I07598; NID:g401771; PIDN:AAC82496.1; PID:g401772  
C:Comment: Although ribosomal protein S6 (see PIR:R3H06) is phosphorylated by this enzyme  
C:Genetics:  
A:Gene: GDB:RPS6KA2; PP90RSK3; RSK; HU-2; RSK3  
A:Cross-references: GDB:365645; OMIM:601685  
A:Map position: 6q27-6q27  
C:Superfamily: ribosomal protein S6 kinase II; protein kinase homology  
C:Keywords: ATP; autophosphorylation; monomer; phosphoprotein; phosphotransferase; serin  
F:57-318/Domain: protein kinase homology <KIN1>  
F:65-73/Region: protein kinase ATP-binding motif  
F:413-672/Domain: protein kinase homology <KIN2>  
F:421-429/Region: protein kinase ATP-binding motif

Query Match 4.2%; Score 8; DB 2; Length 733;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRKS 135  
|||||  
Db 19 RRKRSRKS 26

RESULT 11  
S73210  
ribosomal protein S12, chloroplast - red alga (Porphyra purpurea) chloroplast  
C:Species: chloroplast Porphyra purpurea  
C:Date: 19-Mar-1997 #sequence\_revision 09-May-1997 #text\_change 21-Jan-2000  
C:Accession: S73210  
R:Reith, M.; Munholland, J.

Plant Mol. Biol. Rep. 13, 333-335, 1995  
A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.  
A:Reference number: S73108  
A:Accession: S73210  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-124 <REI>  
A:Cross-references: EMBL:U38804; NID:g1276652; PIDN:AAC08175.1; PID:g1276755  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995  
C:Genetics:  
A:Gene: rps12  
A:Genome: chloroplast  
C:Superfamily: Escherichia coli ribosomal protein S12  
C:Keywords: chloroplast; protein biosynthesis; ribosome  
F:89/Modified site: beta-methylthiospartic acid (Asp) #status predicted

Query Match 3.6%; Score 7; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRKS 134  
|||||  
Db 110 RRKRSRKS 116

RESULT 12  
F89858  
conserved hypothetical protein SA0786 [imported] - Staphylococcus aureus (strain N315  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 17-May-2002  
C:Accession: F89858  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: F89858  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-127 <KUR>  
A:Cross-references: GB:BA000018; PID:g13700728; PIDN:BABA2025.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0786  
C:Superfamily: Bacillus subtilis hypothetical protein yutD

Query Match 3.6%; Score 7; DB 2; Length 127;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105  
|||||  
Db 75 EYCNFGC 81

RESULT 13  
E70416  
hypothetical protein aq\_1343 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C:Accession: E70416  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: E70416  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-160 <AQF>  
A:Cross-references: GB:AE000735; NID:g2983749; PIDN:AAC07319.1; PID:g2983759; GB:AE00  
A:Experimental source: strain VF5

A:Residues: 1-91 <GLA>  
A:Cross-references: GB:NC 003210; PIDN:CAD00480.1; PID:g16411890; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo2402  
C:Superfamily: Bacillus subtilis hypothetical protein yutD

Query Match 5.7%; Score 11; DB 2; Length 91;  
Best Local Similarity 100.0%; Pred. No. 0.00035;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 EYCNFGCAYFV 109  
|||||  
Db 76 EYCNFGCAYFV 86

## RESULT 5

AH1744  
B: subtilis YutD protein homolog lin2501 [imported] - Listeria innocua (strain Clip11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 17-May-2002  
C:Accession: AH1744  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AH1744  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC97728.1; PID:g16415023; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin2501  
C:Superfamily: Bacillus subtilis hypothetical protein yutD

Query Match 5.7%; Score 11; DB 2; Length 91;  
Best Local Similarity 100.0%; Pred. No. 0.00035;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 EYCNFGCAYFV 109  
|||||  
Db 76 EYCNFGCAYFV 86

## RESULT 6

A84079  
hypothetical protein BH3433 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 17-May-2002  
C:Accession: A84079  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira, Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: A84079  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-93 <STO>  
A:Cross-references: GB:AP001518; GB:BA000004; MID:g10175792; PIDN:BA07152.1; GSPDB:GN00178  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH3433  
C:Superfamily: Bacillus subtilis hypothetical protein yutD

Query Match 5.7%; Score 11; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00035;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 EYCNFGCAYFV 109  
|||||  
Db 75 EYCNFGCAYFV 85

## RESULT 7

T23176  
hypothetical protein K01D12.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2000  
C:Accession: T23176  
R:Dobson, R.  
submitted to the EMBL Data Library, June 1996  
A:Reference number: Z19703  
A:Accession: T23176  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-109 <WIL>  
A:Cross-references: EMBL:Z75543; PIDN:CAA99865.1; GSPDB:GN00023; CESP:K01D12.9  
A:Experimental source: clone K01D12  
C:Genetics:  
A:Gene: CESP:K01D12.9  
A:Map position: 5  
A:Introns: 38/1  
C:Superfamily: hypothetical protein K01D12.8

Query Match 4.2%; Score 8; DB 2; Length 109;  
Best Local Similarity 100.0%; Pred. No. 0.73;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 157 SKKRKDTK 164  
|||||  
Db 73 SKKRKDTK 80

## RESULT 8

B86647  
hypothetical protein yb1c [imported] - Lactococcus lactis subsp. lactis (strain IL140)  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: B86647  
R:Bolojin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; EH Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: B86647  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-251 <STO>  
A:Cross-references: GB:AE005176; PID:g12723030; PIDN:AAK04276.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: yb1c

Query Match 4.2%; Score 8; DB 2; Length 251;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LYNKYPG 17  
|||||  
Db 10 LYNKYPG 17

## RESULT 9

S64928  
hypothetical protein YLR094c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein L2534  
C:Species: Saccharomyces cerevisiae  
C:Date: 01-Aug-1995 #sequence\_revision 24-May-1996 #text\_change 19-Apr-2002  
C:Accession: S64928  
R:Benes, V.; Rechmann, S.; Nentwich, U.; Schwager, C.; Ansoorge, W.; Voss, H.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64920

979 5 2.6 119 2 T04243 acidic ribosomal p  
 980 5 2.6 120 2 I34487 T-cell receptor al  
 981 5 2.6 120 2 AG0808 conserved hypothet  
 982 5 2.6 120 2 T44670 V-ATPase proteolip  
 983 5 2.6 120 2 F70966 hypothetical prote  
 984 5 2.6 120 2 S73654 hypothetical prote  
 985 5 2.6 120 2 C82909 hypothetical prote  
 986 5 2.6 120 2 H3761 hypothetical prote  
 987 5 2.6 120 2 T41017 nucleic acid-bind  
 988 5 2.6 121 1 J01646 hypothetical prote  
 989 5 2.6 121 1 H71053 hypothetical prote  
 990 5 2.6 121 2 I27887 Ig heavy chain v r  
 991 5 2.6 121 2 A27888 Ig heavy chain v r  
 992 5 2.6 121 2 H27887 Ig heavy chain v r  
 993 5 2.6 121 2 S40327 Ig kappa chain - h  
 994 5 2.6 121 2 A70443 ribosomal protein  
 995 5 2.6 121 2 D75089 hypothetical prote  
 996 5 2.6 121 2 S48420 probable membrane  
 997 5 2.6 121 2 C72627 hypothetical prote  
 998 5 2.6 121 2 AF0107 hypothetical prote  
 999 5 2.6 122 1 R5DV14 ribosomal protein  
 1000 5 2.6 122 2 B30515 Ig heavy chain v r

## ALIGNMENTS

RESULT 1  
 F70023  
 hypothetical protein yutd - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: F70023  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni, A.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinolis, A.; Laubers, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, T.; Winters, P.; Wipat, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: F70023  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-102 <KUN>  
 A:Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15221.1; PID:g2635728  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yutd  
 C:Superfamily: Bacillus subtilis hypothetical protein yutd

Query Match 6.2%; Score 12; DB 2; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 99 EYCNFGCAVFL 110  
 |||||  
 DB 86 EYCNFGCAVFL 97

RESULT 2  
 H95088  
 conserved hypothetical protein SP0767 [imported] - Streptococcus pneumoniae (strain TIGR  
 C:Species: Streptococcus pneumoniae  
 C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: H95088  
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hoon, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfl, nson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
 A:Reference number: A95000; MUID:21357209; PMID:11463916  
 A:Accession: H95088  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-170 <KUR>  
 A:Cross-references: GB:AF005672; PIDN:AAK74905.1; PID:gl4972242; GSPDB:GN00164; TIGR:  
 A:Experimental source: strain TIGR4  
 C:Genetics:  
 A:Gene: SP0767

Query Match 6.2%; Score 12; DB 2; Length 170;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 69 NEQLRLRGFYKD 80  
 |||||  
 DB 69 NEQLRLRGFYKD 80

RESULT 3  
 C97956  
 hypothetical protein spr0675 [imported] - Streptococcus pneumoniae (strain R6)  
 C:Species: Streptococcus pneumoniae  
 C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
 C:Accession: C97956  
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
 y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: C97956  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-176 <KUR>  
 A:Cross-references: GB:AE007317; PIDN:AAK99479.1; PID:gl5458263; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: spr0675

Query Match 6.2%; Score 12; DB 2; Length 176;  
 Best Local Similarity 100.0%; Pred. No. 5e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 69 NEQLRLRGFYKD 80  
 |||||  
 DB 69 NEQLRLRGFYKD 80

RESULT 4  
 AB1375  
 B. subtilis yutd protein homolog lmo2402 [imported] - Listeria monocytogenes (strain  
 C:Species: Listeria monocytogenes  
 C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 17-May-2002  
 C:Accession: AB1375  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
 .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsthi,  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
 A:Title: Comparative genomics of Listeria species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AB1375  
 A>Status: preliminary  
 A:Molecule type: DNA

833	5	2.6	96	2	AF0741	conserved hypot	906	108	2	AF1935	hypothetical prote
834	5	2.6	96	2	B85797	yebG protein [simi	907	109	1	K3HUSI	Ig kappa chain V-I
835	5	2.6	96	2	F90948	yebG protein [simi	908	109	1	K3HUGO	Ig kappa chain V-I
836	5	2.6	96	2	S76170	hypothetical prote	909	109	1	K3HUTI	Ig kappa chain V-I
837	5	2.6	96	2	C95403	hypothetical prote	910	109	1	K3HUGL	Ig kappa chain V-I
838	5	2.6	96	2	AF0151	probable bacteriop	911	109	2	PH0963	Ig kappa chain V r
839	5	2.6	96	2	AE3640	hypothetical prote	912	109	2	G30601	Ig kappa chain V-I
840	5	2.6	97	2	AC3259	LSU ribosomal prot	913	109	2	B30601	Ig kappa chain V-I
841	5	2.6	98	1	OPCH2	acylphosphatase (E	914	109	2	A30608	Ig kappa chain V-I
842	5	2.6	98	2	G81251	NADH2 dehydrogenas	915	109	2	D30601	Ig kappa chain V-I
843	5	2.6	98	2	S40361	Ig kappa chain V-J	916	109	2	F30607	Ig kappa chain V-I
844	5	2.6	98	2	D72800	gpA protein - Myco	917	109	2	G30601	Ig kappa chain V-I
845	5	2.6	98	2	A46026	group I intron end	918	109	2	G30607	Ig kappa chain V-I
846	5	2.6	99	2	JH0429	Ig kappa chain V-I	919	109	2	H30601	Ig kappa chain V-I
847	5	2.6	99	2	SL12161	polyubiquitin - la	920	109	2	S57098	probable membrane
848	5	2.6	99	2	S62880	ubiquitin/ribosoma	921	109	2	C82913	conserved hypot
849	5	2.6	99	2	H90594	hypothetical prote	922	109	2	H81342	hypothetical prote
850	5	2.6	100	1	K3HUNG	Ig kappa chain pre	923	110	2	S20635	Ig kappa chain V r
851	5	2.6	100	2	T08248	TATA-binding trans	924	110	2	E30607	Ig kappa chain V-I
852	5	2.6	100	2	D71882	hypothetical prote	925	110	2	S44120	Ig kappa chain V-J
853	5	2.6	100	2	G71549	hypothetical prote	926	110	2	A24092	T-cell receptor al
854	5	2.6	100	2	F84766	hypothetical prote	927	110	2	B24092	T-cell receptor al
855	5	2.6	100	2	AC0684	hypothetical prote	928	110	2	S15608	hypothetical prote
856	5	2.6	100	2	AE3219	hypothetical prote	929	110	2	B70343	conserved hypot
857	5	2.6	100	2	S41505	12RNP2 protein - S	930	110	2	JC2327	hypothetical 12.6K
858	5	2.6	100	2	T50611	hypothetical prote	931	110	2	F69760	hypothetical prote
859	5	2.6	100	2	G96948	uncharacterized sm	932	110	2	S15153	hypothetical prote
860	5	2.6	101	2	T43777	ribosomal protein	933	110	2	B70601	hypothetical prote
861	5	2.6	101	2	H84076	hypothetical prote	934	110	2	B97412	hypothetical prote
862	5	2.6	101	2	C72741	hypothetical prote	935	111	2	S40359	Ig kappa chain V-J
863	5	2.6	101	2	A05124	hypothetical prote	936	111	2	E75033	hypothetical prote
864	5	2.6	101	2	D72328	hypothetical prote	937	112	2	A86604	periplasmic divale
865	5	2.6	101	2	S65494	sperm-specific pro	938	112	2	B72019	periplasmic divale
866	5	2.6	101	2	C97704	hypothetical prote	939	112	2	S74214	fmdB protein - Met
867	5	2.6	101	2	AH2395	hypothetical prote	940	112	2	T07633	polyubiquitin 6 -
868	5	2.6	102	2	T01033	hypothetical prote	941	113	2	S26266	T-cell receptor be
869	5	2.6	102	2	S6190	hypothetical prote	942	113	2	PT0741	T-cell receptor be
870	5	2.6	102	2	S65340	hypothetical prote	943	113	2	S17385	T-cell receptor be
871	5	2.6	102	2	T29651	hypothetical prote	944	113	2	S17386	T-cell receptor be
872	5	2.6	103	2	S28823	transcription fact	945	113	2	S26262	T-cell receptor be
873	5	2.6	103	2	PH1045	Ig light chain V r	946	113	2	S26263	T-cell receptor be
874	5	2.6	103	2	B81160	conserved hypot	947	113	2	T44042	chemokine [impor
875	5	2.6	103	2	D52350	conserved hypot	948	113	2	C89828	conserved hypot
876	5	2.6	103	2	G86346	F24J8.2 protein -	949	114	2	S46375	Ig kappa chain V-J
877	5	2.6	104	2	A24424	T-cell receptor be	950	114	2	F83959	ribosomal protein
878	5	2.6	104	2	PH0964	Ig kappa chain V r	951	114	2	B24733	myosin heavy chain
879	5	2.6	104	2	S28596	hypothetical prote	952	114	2	G72611	hypothetical prote
880	5	2.6	105	1	B71334	probable RNA-bind	953	114	2	T09490	hypothetical prote
881	5	2.6	105	2	S11296	polyubiquitin (clo	954	114	2	T17699	arginine-rich prot
882	5	2.6	105	2	T14406	pectin esterase ho	955	114	2	B97036	probable metal-bin
883	5	2.6	105	2	AH2094	RNA-binding protei	956	114	2	C97205	uncharacterized pr
884	5	2.6	105	2	S27493	nodC protein - Bra	957	115	2	JC5908	T cell receptor va
885	5	2.6	105	2	S56482	hypothetical prote	958	115	2	H75543	hypothetical prote
886	5	2.6	105	2	A48943	phage antigenic de	959	115	2	B90263	hypothetical prote
887	5	2.6	106	2	S71947	ubiquitin thiolest	960	116	2	B25521	Ig kappa chain pre
888	5	2.6	106	2	T25815	hypothetical prote	961	116	2	B27594	Ig kappa chain pre
889	5	2.6	106	2	D89551	protein K10C2.6 [i	962	116	2	C27594	Ig kappa chain pre
890	5	2.6	106	2	D97815	hypothetical prote	963	116	2	S62908	ubiquitin precurs
891	5	2.6	106	2	H89982	truncated map-w pr	964	116	2	S24989	ribosomal protein
892	5	2.6	107	1	ERAD51	early E3A 12.5K pr	965	116	2	S50805	hypothetical prote
893	5	2.6	107	2	PH0965	Ig kappa chain V r	966	116	2	S64326	hypothetical prote
894	5	2.6	107	2	S07663	hypothetical prote	967	116	2	C72265	hypothetical prote
895	5	2.6	107	2	S07818	homeotic protein H	968	117	2	I68824	T-cell receptor al
896	5	2.6	108	2	C30608	Ig kappa chain V-I	969	117	2	T24263	hypothetical prote
897	5	2.6	108	2	S33988	Ig kappa chain V r	970	118	2	T03036	Ig light chain - h
898	5	2.6	108	2	E30609	Ig kappa chain V-I	971	118	2	S25001	ubiquitin/ribosoma
899	5	2.6	108	2	A44151	Ig kappa chain V r	972	118	2	C41868	traE1 protein - En
900	5	2.6	108	2	H72224	hypothetical prote	973	118	2	S62422	hypothetical prote
901	5	2.6	108	2	T38079	very hypothetical	974	118	2	S75222	hypothetical prote
902	5	2.6	108	2	S44804	F10E9.5 protein -	975	119	2	E89999	holo-ACP synthase
903	5	2.6	108	2	T49144	hypothetical prote	976	119	2	T07487	hypothetical prote
904	5	2.6	108	2	S74608	hypothetical prote	977	119	2	D83723	hypothetical prote
905	5	2.6	108	2	G69261	hypothetical prote	978	119	2	E81655	conserved hypot



687	5	2.6	64	2	S58141	gene 12 protein -	760	5	2.6	84	2	G64231	modulation protein
688	5	2.6	64	2	T12921	hypothetical prote	761	5	2.6	84	2	S36762	dephosphin long fo
689	5	2.6	64	2	A12881	hypothetical prote	762	5	2.6	84	2	H59106	hypothetical prote
690	5	2.6	65	2	G02999	BRCA1 - rhesus mac	763	5	2.6	84	2	E97807	hypothetical prote
691	5	2.6	66	2	B82818	hypothetical prote	764	5	2.6	85	2	A60622	probable bacteriop
692	5	2.6	67	2	T12871	hypothetical prote	765	5	2.6	85	2	AD2497	hypothetical prote
693	5	2.6	67	2	F97130	probable transcrip	766	5	2.6	86	2	T12077	late embryogenesis
694	5	2.6	67	2	AF2126	hypothetical prote	767	5	2.6	87	1	W4WL51	E4 protein - human
695	5	2.6	68	2	AG3217	hypothetical prote	768	5	2.6	87	2	S16843	Ig kappa chain V r
696	5	2.6	69	2	S01884	ubiquitin / riboso	769	5	2.6	87	2	S16842	Ig kappa chain V r
697	5	2.6	69	2	A64527	hypothetical prote	770	5	2.6	87	2	S16841	Ig kappa chain V r
698	5	2.6	69	2	C71980	hypothetical prote	771	5	2.6	87	2	S16823	Ig kappa chain V r
699	5	2.6	69	2	T03361	gene e6 protein -	772	5	2.6	87	2	S09534	CP75 protein - pha
700	5	2.6	70	2	A48841	dynamlin - bovine (	773	5	2.6	88	2	E97510	hypothetical prote
701	5	2.6	70	2	AD2729	conserved hypothet	774	5	2.6	88	1	UQSY	ubiquitin precursor
702	5	2.6	71	1	S15869	ribosomal protein	775	5	2.6	88	2	T13183	repressor protein
703	5	2.6	71	1	B34209	trichohyalin - gui	776	5	2.6	88	2	B33613	9.8K protein - Ber
704	5	2.6	71	2	F96600	protein F14J16.22	777	5	2.6	88	2	T34429	hypothetical prote
705	5	2.6	71	2	A81037	probable secretion	778	5	2.6	89	2	S34096	Ig kappa chain V r
706	5	2.6	71	2	A09927	probable secretory	779	5	2.6	89	2	A47074	lipopolysaccharide
707	5	2.6	72	2	S34404	ribosomal protein	780	5	2.6	89	2	T35348	probable membrane
708	5	2.6	72	2	A81223	hypothetical prote	781	5	2.6	90	1	S74690	conserved hypothet
709	5	2.6	72	2	T31010	hypothetical prote	782	5	2.6	90	2	S78047	DNA-directed RNA p
710	5	2.6	73	2	S34406	ribosomal protein	783	5	2.6	90	2	E86681	prophage pil prote
711	5	2.6	73	2	A83195	conserved hypothet	784	5	2.6	90	2	T17573	hypothetical prote
712	5	2.6	73	2	T25926	hypothetical prote	785	5	2.6	90	2	T29013	hypothetical prote
713	5	2.6	74	2	E84234	hypothetical prote	786	5	2.6	90	2	A97755	hypothetical prote
714	5	2.6	76	1	UQHU	ubiquitin - human	787	5	2.6	91	2	S67940	Ig kappa chain V r
715	5	2.6	76	1	UQBO	ubiquitin - bovine	788	5	2.6	91	2	S67940	Ig kappa chain V r
716	5	2.6	76	1	UQQA	ubiquitin - oat	789	5	2.6	91	2	E69087	signal recognition
717	5	2.6	76	1	UQUVSF	ubiquitin - fall a	790	5	2.6	91	2	B72200	hypothetical prote
718	5	2.6	76	1	A70028	hypothetical prote	791	5	2.6	92	2	S37524	Ig kappa chain V r
719	5	2.6	76	2	A26087	ubiquitin - fruit	792	5	2.6	92	2	S37513	Ig kappa chain V r
720	5	2.6	76	2	S28203	ubiquitin - rabbit	793	5	2.6	92	2	S37517	Ig kappa chain V r
721	5	2.6	76	2	S42750	polyubiquitin UB2	794	5	2.6	92	2	S37507	Ig kappa chain V r
722	5	2.6	76	2	S12114	polyubiquitin - in	795	5	2.6	92	2	S37502	Ig kappa chain V r
723	5	2.6	76	2	S31653	ubiquitin - Giardia	796	5	2.6	92	2	S37514	Ig kappa chain V r
724	5	2.6	76	2	D69905	ubiquitin - Giardia	797	5	2.6	92	2	S37503	Ig kappa chain V r
725	5	2.6	76	2	H97842	hypothetical prote	798	5	2.6	92	2	S37518	Ig kappa chain V r
726	5	2.6	76	2	A11304	hypothetical prote	799	5	2.6	92	2	T29701	hypothetical prote
727	5	2.6	77	1	UQUT	ubiquitin - Trypan	800	5	2.6	92	2	B97044	hypothetical prote
728	5	2.6	77	1	UQNVAC	ubiquitin - Autogr	801	5	2.6	92	2	C89812	hypothetical prote
729	5	2.6	77	2	S29238	ubiquitin - Entamo	802	5	2.6	93	1	R38219	ribosomal protein
730	5	2.6	77	2	S34334	ubiquitin - Leishm	803	5	2.6	93	1	R32M19	ribosomal protein
731	5	2.6	77	2	S19799	ubiquitin - Entamo	804	5	2.6	93	2	S37508	Ig kappa chain V r
732	5	2.6	77	2	S16263	ubiquitin - potato	805	5	2.6	93	2	S37526	Ig kappa chain V r
733	5	2.6	77	2	C72854	ubiquitin precursor	806	5	2.6	93	2	S37528	Ig kappa chain V r
734	5	2.6	77	2	T41781	viral ubiquitin -	807	5	2.6	93	2	S37505	Ig kappa chain V r
735	5	2.6	77	2	B90849	ubiquitin orf35 -	808	5	2.6	93	2	T10294	ubiquitin-like pro
736	5	2.6	77	2	H85706	hypothetical prote	809	5	2.6	93	2	JQ2029	ubiquitin - Orgyia
737	5	2.6	78	1	UQUTC	unknown protein en	810	5	2.6	93	2	JN0301	ribosomal protein
738	5	2.6	78	1	G69508	ubiquitin precursor	811	5	2.6	93	2	S31184	outer membrane pro
739	5	2.6	78	2	A47243	conserved hypothet	812	5	2.6	93	2	C81783	hypothetical prote
740	5	2.6	78	2	A49039	T-cell receptor be	813	5	2.6	93	2	A69982	hypothetical prote
741	5	2.6	78	2	T03930	T-cell receptor be	814	5	2.6	93	2	A33472	transcription regu
742	5	2.6	78	2	B97027	gene GUT15 protein	815	5	2.6	93	2	S72863	hypothetical prote
743	5	2.6	79	2	D86945	feoa-like protein,	816	5	2.6	93	2	E69163	hypothetical prote
744	5	2.6	79	2	T44149	hypothetical prote	817	5	2.6	94	2	S11248	ubiquitin / riboso
745	5	2.6	79	2	B83400	hypothetical prote	818	5	2.6	94	2	DT4117	transcription regu
746	5	2.6	79	2	B69874	hypothetical prote	819	5	2.6	95	2	JQ0492	ubiquitin 2 - Tetr
747	5	2.6	80	2	S35775	T-cell receptor al	820	5	2.6	95	2	S25154	ubiquitin / riboso
748	5	2.6	80	2	T28305	ORF MSV144 probabl	821	5	2.6	95	2	T41855	AcMNPV orf117 - Bo
749	5	2.6	80	2	E83128	hypothetical prote	822	5	2.6	95	2	G70244	conserved hypothet
750	5	2.6	80	2	I68668	transcription fact	823	5	2.6	95	2	H83360	hypothetical prote
751	5	2.6	80	2	B95851	hypothetical prote	824	5	2.6	95	2	D83237	hypothetical prote
752	5	2.6	82	2	H81138	conserved hypothet	825	5	2.6	95	2	F81873	probable membrane
753	5	2.6	82	2	H82208	hypothetical prote	826	5	2.6	95	2	D64389	hypothetical prote
754	5	2.6	82	2	D64329	hypothetical prote	827	5	2.6	95	2	T40664	very hypothetical
755	5	2.6	82	2	E97781	RP511 protein homo	828	5	2.6	96	1	ASLJSC	protein gp48 - pha
756	5	2.6	83	2	S75314	phycocyanin-associ	829	5	2.6	96	1	T13135	vpr protein - simi
757	5	2.6	83	2	T07252	hypothetical prote	830	5	2.6	96	2	A30601	Ig kappa chain V-I
758	5	2.6	83	2	H89471	protein B0294.2 [i	831	5	2.6	96	2	H70340	hypothetical prote
759	5	2.6	84	2	S01981	ribosomal protein	832	5	2.6	96	2	H64946	yebG protein - Esc

541	6	3.1	983	2	B83055	probable two-compo	614	2052	2	T18290	FYVE finger-contai
542	6	3.1	994	2	B82843	valyl-tRNA synthet	615	2094	2	S33124	tptr protein - huma
543	6	3.1	994	2	S00960	hypothetical prote	616	2172	2	T20145	hypothetical prote
544	6	3.1	1000	2	JE0110	mitotic control pr	617	2364	1	A4159	spectrin beta-G ch
545	6	3.1	1002	2	T43236	carboxypeptidase C	618	2388	1	JE0271	beta spectrin, bet
546	6	3.1	1002	2	T02488	hypothetical prote	619	2458	2	T17420	probable polyketid
547	6	3.1	1007	2	T32285	hypothetical prote	620	2774	2	A43359	microtubule-associ
548	6	3.1	1011	2	S11177	vacuolar sorting p	621	2897	2	B48666	cell proliferation
549	6	3.1	1011	2	T20785	hypothetical prote	622	3295	2	A48666	cell proliferation
550	6	3.1	1032	2	A11697	alpha-mannosidase	623	3295	2	AE0074	probable adhesin Y
551	6	3.1	1033	2	S19247	cell adhesion prot	624	6658	2	T13931	projectin - fruit
552	6	3.1	1036	2	AG1326	alpha-mannosidase	625	7576	2	T17428	FK506 polyketide s
553	6	3.1	1039	2	S18199	myosin heavy chain	626	8243	2	T31307	type I fatty acid
554	6	3.1	1053	2	S34172	sucrose-phosphate	627	15	2	S10891	ubiquitin thiolest
555	6	3.1	1057	1	OYRTR	atrial natriuretic	628	15	2	PT0093	ubiquitin thiolest
556	6	3.1	1057	1	OYMSAR	atrial natriuretic	629	20	2	S28405	ubiquitin thiolest
557	6	3.1	1057	2	I57963	natriuretic peptid	630	20	2	S27350	lamin B receptor -
558	6	3.1	1057	2	I55319	guanylyl cyclase A	631	22	2	S19802	lysophospholipase
559	6	3.1	1061	1	OYH0AR	natriuretic peptid	632	23	2	S45031	ubiquitin - potato
560	6	3.1	1072	2	T18802	hypothetical prote	633	25	2	S17561	homeotic protein S
561	6	3.1	1073	2	D87596	TonB-dependent rec	634	27	2	B47295	ubiquitin thiolest
562	6	3.1	1084	2	S57681	CRM1 protein - yea	635	30	2	PD0006	homeodomain protei
563	6	3.1	1091	2	S57112	JSN1 protein - yea	636	34	2	H30607	cysteine synthase
564	6	3.1	1113	2	S28925	nuclear pore compl	637	34	2	J21080	ig kappa chain V-I
565	6	3.1	1118	2	T27865	hypothetical prote	638	36	2	SC2006	sperm protein EM1
566	6	3.1	1159	2	A34308	Ca2+-transporting	639	36	2	A69827	differentiation in
567	6	3.1	1160	2	T31688	Ca2+-transporting	640	38	2	E49164	hypothetical prote
568	6	3.1	1163	2	E71481	probable swf/snf h	641	38	2	C83729	chromogranin-B - r
569	6	3.1	1166	2	G86596	SWI/SNF family hel	642	41	2	A45942	hypothetical prote
570	6	3.1	1166	2	G72027	swi/snf family hel	643	41	2	T45906	myosin heavy chain
571	6	3.1	1175	2	H83437	hypothetical prote	644	42	2	S20863	hypothetical prote
572	6	3.1	1188	2	G72734	hypothetical prote	645	42	2	JH0383	ubiquitin - Norway
573	6	3.1	1210	2	I39410	AF-4 protein, spli	646	42	2	JH0384	hypothetical 5.1K
574	6	3.1	1213	2	A58198	serine/proline-ric	647	42	2	I65249	hypothetical 5.1K
575	6	3.1	1217	2	T00607	hypothetical prote	648	42	2	T07324	homeoprotein 13A -
576	6	3.1	1230	2	T22458	hypothetical prote	649	44	2	H90641	hypothetical prote
577	6	3.1	1272	2	C90593	hypothetical prote	650	44	2	H85492	hypothetical prote
578	6	3.1	1278	2	B70236	hypothetical prote	651	45	2	C82451	hypothetical prote
579	6	3.1	1333	2	B84601	probable retroelem	652	46	2	PD0024	protein kinase (EC
580	6	3.1	1335	2	H75511	DNA polymerase III	653	50	2	B64035	hypothetical prote
581	6	3.1	1338	2	T02206	hypothetical prote	654	51	2	S26101	outer membrane pro
582	6	3.1	1344	2	T42637	hypothetical prote	655	51	2	T07326	hypothetical prote
583	6	3.1	1366	2	T35965	probable large pro	656	53	2	A56512	hypothetical prote
584	6	3.1	1403	2	S64142	hypothetical prote	657	54	2	S40381	nuclear pore prote
585	6	3.1	1404	2	T19277	hypothetical prote	658	54	2	A25521	ig kappa chain V-I
586	6	3.1	1407	1	S28589	trichohyalin - rab	659	55	1	R5NT32	ig kappa chain V r
587	6	3.1	1444	2	T18856	angiogenesis inhib	660	55	2	S43121	ribosomal protein
588	6	3.1	1455	2	A70199	hypothetical prote	661	55	2	T06496	ubiquitin - basidi
589	6	3.1	1477	2	S64616	YOR1 protein - yea	662	55	2	A28911	ubiquitin/ribosoma
590	6	3.1	1489	2	S60416	DNA helicase YGL15	663	55	2	T09464	male accessory gla
591	6	3.1	1491	1	S43793	copper-transportin	664	55	2	E48362	H+-transporting tw
592	6	3.1	1496	2	I51170	carbamyl phosphate	665	56	2	S45027	hypothetical prote
593	6	3.1	1502	2	T48309	hypothetical prote	666	56	2	G90278	sox1 protein - Her
594	6	3.1	1512	2	G85090	hypothetical prote	667	56	2	C82226	hypothetical prote
595	6	3.1	1519	2	G71402	DNA (cytosine-5)-	668	56	2	AC1987	hypothetical prote
596	6	3.1	1547	2	T13847	sno protein - frui	669	57	2	G84072	hypothetical prote
597	6	3.1	1549	1	A40691	trichohyalin - she	670	57	2	G97083	hypothetical prote
598	6	3.1	1564	2	S55517	probable transport	671	58	2	B48663	pyruvate kinase (E
599	6	3.1	1592	2	T16055	hypothetical prote	672	59	2	D81084	hypothetical prote
600	6	3.1	1601	2	AE2011	hypothetical prote	673	60	2	S35929	hypothetical prote
601	6	3.1	1605	2	T31435	DNA-directed RNA p	674	60	2	JT0491	pyruvate kinase (E
602	6	3.1	1653	2	A36349	clathrin heavy cha	675	60	2	S09227	ubiquitin 1 - Tetr
603	6	3.1	1696	2	T27447	hypothetical prote	676	60	2	B69189	homeotic protein H
604	6	3.1	1726	2	A39401	merozoite surface	677	62	2	AH1301	conserved hypothet
605	6	3.1	1731	2	A45604	major blood-stage	678	62	2	AH1673	ribosomal protein
606	6	3.1	1755	2	F82618	chemotaxis-related	679	62	2	T10302	ribosomal protein
607	6	3.1	1778	2	T50074	probable nucleopor	680	62	2	C97621	hypothetical prote
608	6	3.1	1822	2	S63985	collagen alpha 2 c	681	62	2	AI2130	hypothetical prote
609	6	3.1	1852	1	VJCH2	vitellogenin II pr	682	63	2	B30606	T-cell receptor al
610	6	3.1	1898	1	A45973	trichohyalin - hum	683	63	2	I48071	GTP binding protei
611	6	3.1	1906	2	AD2443	hypothetical prote	684	63	2	G96517	protein T2E6.24 [l
612	6	3.1	1939	2	I48175	myosin heavy chain	685	63	2	F84501	10-formyltetrahydr
613	6	3.1	1944	2	T40065	tRNA-splicing endo	686	64	2	S17384	T-cell receptor be

395	6	3.1	550	1	E28544	methyl coenzyme M	468	6	3.1	728	2	D86278	hypothetical prote
396	6	3.1	550	2	B69022	methyl coenzyme M	469	6	3.1	728	2	A54603	transcription fact
397	6	3.1	552	2	C96563	probable protein k	470	6	3.1	732	2	T18567	hypothetical prote
398	6	3.1	553	2	T14354	probable somatic e	471	6	3.1	749	2	T34090	hypothetical prote
399	6	3.1	554	2	T31783	hypothetical prote	472	6	3.1	752	2	D95246	glycogen phosphory
400	6	3.1	557	2	A56690	esterase - Caenorh	473	6	3.1	752	2	A98111	maltoextrin phosph
401	6	3.1	557	2	E72282	oligopeptide ABC t	474	6	3.1	752	2	T16508	hypothetical prote
402	6	3.1	560	2	S58741	probable site-spec	475	6	3.1	752	2	T26508	hypothetical prote
403	6	3.1	560	2	T40608	conserved hypotet	476	6	3.1	754	1	JC5314	CDC38/cdc2-like Ki
404	6	3.1	562	2	T42250	polypeptide N-acet	477	6	3.1	754	1	AF2849	methyl-accepting c
405	6	3.1	562	2	T26242	hypothetical prote	478	6	3.1	761	2	T15776	hypothetical prote
406	6	3.1	563	2	T32504	hypothetical prote	479	6	3.1	765	2	T23582	3',5'-cyclic-nucie
407	6	3.1	564	2	T28949	hypothetical prote	480	6	3.1	768	2	T10796	hypothetical prote
408	6	3.1	566	2	T50536	succinate dehydrog	481	6	3.1	768	2	D97626	hypothetical prote
409	6	3.1	566	2	T26243	hypothetical prote	482	6	3.1	773	2	T00502	probable receptor-
410	6	3.1	568	2	S15008	hypothetical prote	483	6	3.1	779	2	E97778	endopeptidase La (
411	6	3.1	570	2	E29525	gene disco protein	484	6	3.1	784	2	AC0383	endopeptidase La (
412	6	3.1	570	2	G81380	probable AAA famil	485	6	3.1	790	2	S50082	nuclear cap bindin
413	6	3.1	571	2	F84718	hypothetical prote	486	6	3.1	791	2	E90566	conserved hypotet
414	6	3.1	573	2	E84687	hypothetical prote	487	6	3.1	796	2	E90566	hypothetical prote
415	6	3.1	573	2	H96744	probable cytosolic	488	6	3.1	803	2	T40514	Chaperonin hsp78p
416	6	3.1	574	2	T04249	hypothetical prote	489	6	3.1	807	2	T06409	valosin-containing
417	6	3.1	576	2	S22127	protein kinase pol	490	6	3.1	809	1	S60112	transitional endop
418	6	3.1	577	2	T04229	ABC-type transport	491	6	3.1	813	2	C97292	ATPases with chape
419	6	3.1	577	2	S47961	glucosyltransferas	492	6	3.1	815	2	T46169	CDC48-like protein
420	6	3.1	580	2	AH2138	ABC transporter Ar	493	6	3.1	815	2	B56708	mannose-1-phosphat
421	6	3.1	581	2	T05232	hypothetical prote	494	6	3.1	815	2	G97266	probable membrane
422	6	3.1	583	2	T25690	hypothetical prote	495	6	3.1	817	2	H75035	DNA-directed RNA p
423	6	3.1	585	2	C82157	hypothetical prote	496	6	3.1	822	2	S26872	1,4-beta-D-glucan
424	6	3.1	586	2	T49210	hypothetical prote	497	6	3.1	823	2	B87348	ATP-dependent heli
425	6	3.1	588	2	E96533	probable Serine/Th	498	6	3.1	824	2	D64738	ATP-dependent heli
426	6	3.1	592	2	T48346	photoreceptor-inte	499	6	3.1	824	2	H90647	helicase, ATP-depe
427	6	3.1	592	2	F86242	unknown protein, 9	500	6	3.1	824	2	H85498	potassium transpor
428	6	3.1	600	2	C83142	hypothetical prote	501	6	3.1	826	2	T02268	hypothetical prote
429	6	3.1	601	2	H96740	hypothetical prote	502	6	3.1	826	2	T23072	DNA gyrase, sub A
430	6	3.1	603	2	B64444	hypothetical prote	503	6	3.1	827	2	E64607	hypothetical prote
431	6	3.1	603	2	B64444	hypothetical prote	504	6	3.1	832	2	T33490	glycerol-3-phospha
432	6	3.1	607	2	B86453	F9J11.4 protein -	505	6	3.1	834	2	F83185	transitional endop
433	6	3.1	613	2	D86147	TiN6.3 protein - A	506	6	3.1	835	1	S67669	transitional endop
434	6	3.1	617	2	T42249	polypeptide N-acet	507	6	3.1	843	2	T48355	cation-transportin
435	6	3.1	618	2	T42248	polypeptide N-acet	508	6	3.1	847	2	G75270	probable receptor
436	6	3.1	619	2	F70480	conserved hypotet	509	6	3.1	852	2	A85041	proliferating-cell
437	6	3.1	620	2	S48772	AFR1 protein - yea	510	6	3.1	855	2	A48168	3',5'-cyclic-GMP p
438	6	3.1	620	2	T11221	DNA polymerase hom	511	6	3.1	859	1	S08418	3',5'-cyclic-GMP p
439	6	3.1	621	2	I57939	taurine transporte	512	6	3.1	859	2	B34611	3',5'-cyclic-GMP p
440	6	3.1	621	2	F70334	hypothetical prote	513	6	3.1	859	2	S13030	hypothetical prote
441	6	3.1	622	2	AF1074	lytic murein trans	514	6	3.1	859	2	S64952	lipoxigenase (EC 1
442	6	3.1	628	2	D86466	69.4K hypothetical	515	6	3.1	862	2	T07775	phycobilisome anch
443	6	3.1	629	2	T24469	hypothetical prote	516	6	3.1	863	2	S37088	hypothetical prote
444	6	3.1	634	2	T27959	hypothetical prote	517	6	3.1	863	2	T27958	hypothetical prote
445	6	3.1	639	2	AB3456	long-chain-fatty-a	518	6	3.1	864	2	S60441	hypothetical prote
446	6	3.1	639	2	B24810	beta-conglycinin a	519	6	3.1	870	2	T30110	hypothetical prote
447	6	3.1	639	2	A99225	mRNA 3'-end proces	520	6	3.1	871	2	T43427	pob1 protein - fis
448	6	3.1	640	2	C81108	ABC transporter, A	521	6	3.1	875	2	S75377	probable helicase
449	6	3.1	642	2	S31285	RNA polymerase II	522	6	3.1	881	2	C97362	protein sensor pro
450	6	3.1	647	2	T00370	hypothetical prote	523	6	3.1	881	2	AD3580	two component sens
451	6	3.1	649	2	G71283	probable ATP-depen	524	6	3.1	882	2	S23019	DNA-directed DNA p
452	6	3.1	650	2	T21498	hypothetical prote	525	6	3.1	882	2	F90201	DNA polymerase I (
453	6	3.1	653	2	S67035	probable membrane	526	6	3.1	892	2	T27005	hypothetical prote
454	6	3.1	655	2	A46270	Na and Cl dependen	527	6	3.1	899	2	S49634	hypothetical prote
455	6	3.1	659	2	T41975	helicase-primase c	528	6	3.1	900	2	F71426	hypothetical prote
456	6	3.1	662	2	E84725	ankyrin-like prote	529	6	3.1	904	2	T22457	hypothetical prote
457	6	3.1	677	2	T39713	zinc finger protel	530	6	3.1	906	1	S32607	trifunctional enzy
458	6	3.1	689	2	S59797	hypothetical prote	531	6	3.1	915	1	A55144	autotaxin precurs
459	6	3.1	692	2	S48316	probable membrane	532	6	3.1	923	2	T08033	probable serine/th
460	6	3.1	693	2	B84647	hypothetical prote	533	6	3.1	924	1	JC1269	DNA-directed DNA p
461	6	3.1	707	2	T29854	spectrin beta chai	534	6	3.1	925	2	C84538	probable LRR recep
462	6	3.1	708	2	T00205	hypothetical prote	535	6	3.1	927	2	T08034	probable serine/th
463	6	3.1	710	2	S67098	probable membrane	536	6	3.1	941	2	F71332	probable chromosom
464	6	3.1	712	2	AG0419	ribonucleoside-tri	537	6	3.1	950	2	G83167	valyl-tRNA synthet
465	6	3.1	713	2	T21201	hypothetical prote	538	6	3.1	960	1	A39651	discs-large tumor
466	6	3.1	714	2	T22454	hypothetical prote	539	6	3.1	962	2	T22459	hypothetical prote
467	6	3.1	722	2	AD1796	hypothetical membr	540	6	3.1	969	2	H69425	modification methy
467	6	3.1	727	2	G01792	transcription fact							

249	3.1	354	2	E69499	iron-sulfur bindin	322	6	3.1	444	2	B96498	probable polygalac
250	3.1	355	1	RGHUI2	GTP-binding regula	323	6	3.1	444	2	C96498	probable polygalac
251	3.1	355	1	RGRTI2	GTP-binding regula	324	6	3.1	444	2	S54039	hypothetical prote
252	3.1	355	1	RGMSI2	GTP-binding regula	325	6	3.1	447	2	S52391	serotomisin B - mo
253	3.1	355	2	A61031	GTP-binding regula	326	6	3.1	449	2	S02011	tricylglycerol li
254	3.1	355	2	I50238	Gi2 protein alpha-	327	6	3.1	449	2	JQ1277	hypothetical prote
255	3.1	355	2	S28158	GTP-binding regula	328	6	3.1	450	2	T16220	hypothetical prote
256	3.1	355	2	T19016	hypothetical prote	329	6	3.1	451	2	T16418	hypothetical prote
257	3.1	358	2	T37805	hypothetical prote	330	6	3.1	452	2	F90452	conserved hypotet
258	3.1	360	2	T43254	GTP-binding protei	331	6	3.1	454	2	F97918	aspartate kinase I
259	3.1	360	2	D69823	flagellar biosynth	332	6	3.1	454	2	G95047	aspartate kinase I
260	3.1	360	2	S39916	cwlI protein - Bac	333	6	3.1	457	1	A28468	chromogranin A pre
261	3.1	361	2	H86763	thiamin biosynthes	334	6	3.1	457	2	A95940	probable integral
262	3.1	362	2	E71637	hypothetical prote	335	6	3.1	457	2	T40770	coiled coil protei
263	3.1	362	2	G97849	hypothetical prote	336	6	3.1	460	2	A96555	unknown protein Ii
264	3.1	363	2	AH2721	fatty acid/phospho	337	6	3.1	463	2	A64381	NADH oxidase - Met
265	3.1	370	1	BVLVWX	mbpX protein - liv	338	6	3.1	464	2	AD2289	hypothetical prote
266	3.1	371	2	B49515	phosphoprotein 75	339	6	3.1	465	1	JC1318	triacylglycerol li
267	3.1	373	2	B90512	DNA polymerase III	340	6	3.1	466	2	T41076	hypothetical prote
268	3.1	374	2	H83693	carboxypeptidase G	341	6	3.1	467	2	T39070	probable ubiquitin
269	3.1	375	2	T09374	hypothetical prote	342	6	3.1	469	2	B82446	GGDEF family prote
270	3.1	378	2	H96773	hypothetical prote	343	6	3.1	471	2	S02016	U1 snRNP 70K prote
271	3.1	379	1	E71162	hypothetical prote	344	6	3.1	472	2	T29313	hypothetical prote
272	3.1	379	2	A90337	proteinase, probab	345	6	3.1	472	2	T25949	hypothetical prote
273	3.1	381	2	T24256	hypothetical prote	346	6	3.1	475	2	A81735	cysteinyl-tRNA syn
274	3.1	382	2	D75001	hypothetical prote	347	6	3.1	476	2	A43942	lipase - Pseudomon
275	3.1	383	2	H71848	probable na+/ht an	348	6	3.1	478	2	S57203	hexokinase - yeast
276	3.1	383	2	G64667	NA+/H+ antiporter	349	6	3.1	481	2	A56346	transcription fact
277	3.1	385	2	JC2256	aspartate transami	350	6	3.1	483	2	D87752	protein C18E3.6 Ii
278	3.1	386	1	S41691	ubiquinol-cytochro	351	6	3.1	483	2	T15180	hypothetical prote
279	3.1	388	2	T26977	hypothetical prote	352	6	3.1	485	1	K1BYHA	hexokinase (EC 2.7
280	3.1	389	2	T34766	iron-sulfur cofact	353	6	3.1	485	2	T17604	hypothetical prote
281	3.1	391	2	S22579	translation initia	354	6	3.1	485	2	D84475	probable Athila re
282	3.1	391	2	T51737	RNA helicase RH2 I	355	6	3.1	487	2	T34887	probable transposa
283	3.1	391	2	T23292	hypothetical prote	356	6	3.1	487	2	B90276	methylmalonate-sem
284	3.1	393	1	A48357	nonstructural prot	357	6	3.1	492	2	T20368	hypothetical prote
285	3.1	395	2	G82298	tyrosyl-tRNA synth	358	6	3.1	493	2	A12314	hypothetical prote
286	3.1	396	2	E97947	cell division prot	359	6	3.1	494	2	A48133	pre-mRNA splicing
287	3.1	399	2	C83563	tyrosyl-tRNA synth	360	6	3.1	496	2	JH0206	hypothetical 57.4K
288	3.1	399	2	F89978	conserved hypotet	361	6	3.1	496	2	T39478	repR protein - Str
289	3.1	399	2	B95080	cell division prot	362	6	3.1	496	2	S06925	replication initia
290	3.1	400	2	A96596	tyrosyl-tRNA synth	363	6	3.1	496	2	B96674	hypothetical prote
291	3.1	402	2	S48770	probable membrane	364	6	3.1	497	2	D72213	conserved hypotet
292	3.1	403	2	B72778	probable processin	365	6	3.1	498	2	D83994	xylose kinase xylB
293	3.1	403	2	T36078	hypothetical prote	366	6	3.1	501	2	S16711	ABC1 protein precu
294	3.1	404	2	A96498	probable polygalac	367	6	3.1	501	2	T21973	hypothetical prote
295	3.1	404	2	H87635	Rieske 2Fe-2S fami	368	6	3.1	502	2	T29729	hypothetical prote
296	3.1	406	2	F90935	acetylornithine de	369	6	3.1	503	2	F82275	toxin co-regulated
297	3.1	406	2	B85784	succinylornithine de	370	6	3.1	503	2	JN0524	hypothetical prote
298	3.1	406	2	D64934	tyrosylornithine	371	6	3.1	504	2	T31784	pyruvate kinase (E
299	3.1	407	1	A55515	translation initia	372	6	3.1	508	2	H64223	protein F14N23.31
300	3.1	407	2	T06824	MSHA biogenesis pr	373	6	3.1	512	2	A86238	probable structure
301	3.1	407	2	C82327	hypothetical prote	374	6	3.1	512	2	T40576	E2 protein - human
302	3.1	410	2	E72397	hypothetical prote	375	6	3.1	514	1	W2WLB5	E2 protein - human
303	3.1	411	2	T25223	3-oxoacyl-l-acyl-ca	376	6	3.1	515	2	F88618	protein W06F12.2 I
304	3.1	416	2	AH2223	MYB DNA-binding-li	377	6	3.1	515	2	JS0083	cellulose 1,4-beta
305	3.1	423	2	T51794	transcription init	378	6	3.1	516	2	S33164	cellulose 1,4-beta
306	3.1	424	2	T14358	transcription init	379	6	3.1	516	2	S60661	legumin - Weiwitsc
307	3.1	424	2	T14359	hypothetical prote	380	6	3.1	519	2	S35575	myosin heavy chain
308	3.1	425	2	T43772	multifunctional pu	381	6	3.1	520	2	G84510	probable receptor-
309	3.1	425	2	S14147	SH3-domain binding	382	6	3.1	520	2	T52584	diacylglycerol O-a
310	3.1	425	2	JE0086	aminoimidazole rib	383	6	3.1	520	2	G84864	probable membrane
311	3.1	425	2	S55684	5-aminoimidazole r	384	6	3.1	521	2	T05994	protein kinase hom
312	3.1	426	2	A35641	probable chromatin	385	6	3.1	523	2	H84510	probable receptor-
313	3.1	431	2	T38471	hypothetical prote	386	6	3.1	524	2	S38794	cellulose 1,4-beta
314	3.1	431	2	T33287	hypothetical prote	387	6	3.1	525	1	T49212	hypothetical prote
315	3.1	438	2	T47711	ankyrin-repeat pro	388	6	3.1	527	2	T10388	hypothetical prote
316	3.1	439	2	JQ1729	diacylglycerol O-a	389	6	3.1	530	2	H72864	AcOrf-119 protein
317	3.1	441	2	H84576	hypothetical prote	390	6	3.1	542	2	D87185	probable UDP-gluc
318	3.1	441	2	T34401	hypothetical prote	391	6	3.1	545	2	AI2327	hypothetical prote
319	3.1	443	2	S37915	CBS domain protein	392	6	3.1	548	2	A56353	MIDI protein - yea
320	3.1	443	2	E95229	hemolysin [importe	393	6	3.1	549	2	D87000	acetyl/propionyl C
321	3.1	443	2	H98093		394	6	3.1				

103	6	3.1	168	2	B97776	hypothetical prote	176	6	3.1	292	2	G88448	protein C45G9.4 [i
104	6	3.1	170	2	D86522	CT144 hypothetical	177	6	3.1	294	2	A44856	haloacetate dehalo
105	6	3.1	171	2	T48393	hypothetical prote	178	6	3.1	294	2	H96662	hypothetical prote
106	6	3.1	173	2	G81428	hypothetical prote	179	6	3.1	294	2	E88448	protein C45G9.9 [i
107	6	3.1	174	2	A11979	hypothetical prote	180	6	3.1	296	2	S60251	mab-18 protein (tr
108	6	3.1	180	2	A81287	ribosomal-protein-	181	6	3.1	296	2	AG3515	N-formylglutamate
109	6	3.1	181	2	T48241	hypothetical prote	182	6	3.1	297	2	B58460	phosphoprotein NS
110	6	3.1	187	1	A45113	ribosomal protein	183	6	3.1	298	2	A70238	hypothetical prote
111	6	3.1	189	2	E90157	conserved hypothet	184	6	3.1	299	2	A13184	transcription regu
112	6	3.1	191	2	D64418	hypothetical prote	185	6	3.1	299	2	E84262	rhannosyl transfer
113	6	3.1	196	2	C95114	exonuclease [impr	186	6	3.1	303	1	TVHUUD	transforming prote
114	6	3.1	197	2	T33525	hypothetical prote	187	6	3.1	303	2	T15694	hypothetical prote
115	6	3.1	198	2	T71181	hypothetical prote	188	6	3.1	303	2	T23583	hypothetical prote
116	6	3.1	201	2	T42647	hypothetical prote	189	6	3.1	304	2	T27583	hypothetical prote
117	6	3.1	202	1	R31V4	ribosomal protein	190	6	3.1	305	2	B72691	probable fumarate
118	6	3.1	202	2	AB2378	hypothetical prote	191	6	3.1	306	2	F87346	conserved hypothet
119	6	3.1	203	2	B70254	conserved hypothet	192	6	3.1	307	2	T47676	zinc finger protei
120	6	3.1	205	2	S34919	ribosomal protein	193	6	3.1	309	2	S77170	site-specific DNA-
121	6	3.1	206	1	E64317	hypothetical prote	194	6	3.1	310	2	E90231	UDP-glucose 4-epim
122	6	3.1	206	2	T36820	hypothetical prote	195	6	3.1	312	2	F86274	hypothetical prote
123	6	3.1	207	2	T24858	hypothetical prote	196	6	3.1	314	2	T46743	carbamate kinase (
124	6	3.1	208	2	H97983	DNA-directed DNA p	197	6	3.1	314	2	AC1484	B. subtilis Yqfa P
125	6	3.1	210	2	A69509	thiamin phosphate	198	6	3.1	314	2	A11123	B. subtilis Yqfa P
126	6	3.1	214	2	T52495	hypothetical prote	199	6	3.1	315	2	H82512	probable 3-hydroxy
127	6	3.1	221	2	H97021	hypothetical prote	200	6	3.1	315	2	T05512	hypothetical prote
128	6	3.1	222	2	S26994	cellulose 1,4-beta	201	6	3.1	317	2	P00339	multimeric flavodo
129	6	3.1	224	2	C81417	probable transcrip	202	6	3.1	317	2	E97211	hypothetical prote
130	6	3.1	226	2	T29210	hypothetical prote	203	6	3.1	318	2	G96910	D-3 phosphoglycera
131	6	3.1	227	2	C72559	hypothetical prote	204	6	3.1	319	2	A69756	adhesion protein h
132	6	3.1	228	2	T23994	hypothetical prote	205	6	3.1	321	2	G59520	probable epimerase
133	6	3.1	230	2	E86765	hypothetical prote	206	6	3.1	321	2	G84750	probable homeodoma
134	6	3.1	232	2	T21526	hypothetical prote	207	6	3.1	321	2	T16451	hypothetical prote
135	6	3.1	236	2	E70028	conserved hypothet	208	6	3.1	323	2	S20099	transforming prote
136	6	3.1	239	2	F75111	hypothetical prote	209	6	3.1	325	2	AB1123	B. subtilis IolC P
137	6	3.1	242	2	A69026	conserved hypothet	210	6	3.1	325	2	AD1483	B. subtilis IolC P
138	6	3.1	243	2	AF1426	reductases homolo	211	6	3.1	326	2	T38861	hypothetical prote
139	6	3.1	244	2	B71022	hypothetical prote	212	6	3.1	327	2	JC1437	glucan endo-1,3-be
140	6	3.1	248	2	H86285	protein F911.13 [i	213	6	3.1	328	2	E72347	flagellar motor sw
141	6	3.1	248	2	A49515	phosphoprotein 41	214	6	3.1	329	2	G98165	hypothetical prote
142	6	3.1	249	2	S07917	RF3 protein - yeas	215	6	3.1	329	2	AH3121	conserved hypothet
143	6	3.1	251	2	T22839	hypothetical prote	216	6	3.1	330	2	S58255	hypothetical prote
144	6	3.1	252	2	F70226	conserved hypothet	217	6	3.1	331	2	T13145	cytosine-specific
145	6	3.1	252	2	H90191	hypothetical prote	218	6	3.1	331	2	C84515	hypothetical prote
146	6	3.1	259	2	F71561	probable rRNA meth	219	6	3.1	332	1	JS0274	biotin synthase (E
147	6	3.1	261	2	B86643	carbonyl reductase	220	6	3.1	332	2	T32083	hypothetical prote
148	6	3.1	261	2	S43218	hypothetical prote	221	6	3.1	335	2	B86594	yop translocation
149	6	3.1	261	2	T22120	hypothetical prote	222	6	3.1	335	2	F72030	type III secretion
150	6	3.1	262	2	G81954	cytochrome C1 prec	223	6	3.1	338	2	T47920	hypothetical prote
151	6	3.1	262	2	F97151	ABC transported MD	224	6	3.1	341	1	TWNSJD	transforming prote
152	6	3.1	264	2	T50640	zinc finger protei	225	6	3.1	341	2	JC4051	jun-D protein - ra
153	6	3.1	265	2	A64705	hypothetical prote	226	6	3.1	343	2	T24700	hypothetical prote
154	6	3.1	265	2	A71815	hypothetical prote	227	6	3.1	344	2	A69661	transcription regu
155	6	3.1	265	2	S62363	D-2-chloropropioni	228	6	3.1	344	2	F69272	hypothetical prote
156	6	3.1	265	2	T43123	probable positive	229	6	3.1	345	2	T41473	probable zinc fing
157	6	3.1	266	2	D81011	ubiquinol-cytochro	230	6	3.1	345	2	C75207	DNA primase FAB223
158	6	3.1	266	2	T49052	hypothetical prote	231	6	3.1	345	2	C89882	conserved hypothet
159	6	3.1	267	2	A13458	urease accessory p	232	6	3.1	346	2	G96643	hypothetical prote
160	6	3.1	268	2	G96016	probable gluconate	233	6	3.1	347	2	A43815	transforming prote
161	6	3.1	270	2	D69127	hypothetical prote	234	6	3.1	349	2	AB0783	probable membrane
162	6	3.1	271	1	IKECM	colicin M - Escher	235	6	3.1	350	2	AG1721	uroporphyrinogen I
163	6	3.1	272	2	AC2037	hypothetical prote	236	6	3.1	351	2	AC3090	conserved hypothet
164	6	3.1	273	2	T18809	hypothetical prote	237	6	3.1	351	2	G98196	hypothetical prote
165	6	3.1	276	2	S70815	glycosyl transfera	238	6	3.1	352	2	T06747	hypothetical prote
166	6	3.1	277	2	T27601	hypothetical prote	239	6	3.1	353	2	F75562	ornithine carbamoy
167	6	3.1	277	2	T40033	probable mitochond	240	6	3.1	353	2	AD1351	uroporphyrinogen I
168	6	3.1	279	2	H83836	hypothetical prote	241	6	3.1	353	2	E97503	fatty acid/phospho
169	6	3.1	280	1	F64440	conserved hypothet	242	6	3.1	354	1	RGH011	GTP-binding regula
170	6	3.1	280	2	A81027	lacto-N-neotetraos	243	6	3.1	354	1	RGBO11	GTP-binding regula
171	6	3.1	280	2	T09576	Ibx1 transcription	244	6	3.1	354	1	RGXT11	GTP-binding regula
172	6	3.1	283	2	T10228	hypothetical prote	245	6	3.1	354	1	RGXLI1	GTP-binding regula
173	6	3.1	286	2	C72456	hypothetical prote	246	6	3.1	354	2	I50237	GTP-binding regula
174	6	3.1	288	2	D84616	probable DHLH tran	247	6	3.1	354	2	S28157	GTP-binding regula
175	6	3.1	292	2	C69106	glucose-1-phosphat	248	6	3.1	354	2	T05975	probable carboxy-t

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: February 26, 2003, 09:58:43 ; Search time 46 Seconds  
(without alignments)  
401.257 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 192

Sequence: 1 MKRVTPEMLNKNYKPGQF.....EQQDKEMTSKQHLLEVRKN 192

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	6.2	102	2 F70023	hypothetical prote
2	12	6.2	170	2 H95088	conserved hypotet
3	12	6.2	176	2 C97956	hypothetical prote
4	11	5.7	91	2 AB1375	B. subtilis YutD p
5	11	5.7	91	2 AH1744	B. subtilis YutD p
6	11	5.7	93	2 A84079	hypothetical prote
7	8	4.2	109	2 T23176	hypothetical prote
8	8	4.2	251	2 B86647	hypothetical prote
9	8	4.2	502	2 S64928	hypothetical prote
10	8	4.2	733	2 A57459	ribosomal protein
11	7	3.6	124	2 S73210	ribosomal protein
12	7	3.6	127	2 F89858	conserved hypotet
13	7	3.6	160	2 E70416	hypothetical prote
14	7	3.6	194	2 E71407	hypothetical prote
15	7	3.6	194	2 T05238	hypothetical prote
16	7	3.6	250	2 T47871	hypothetical prote
17	7	3.6	317	2 JE0175	frezzled protein-1
18	7	3.6	344	2 S59043	spilling factor SR
19	7	3.6	364	2 AC3224	hypothetical prote
20	7	3.6	381	2 I39939	phenylalanine dehy
21	7	3.6	382	1 VH1HPC	nucleocapsid prote
22	7	3.6	382	1 S24282	nucleocapsid prote
23	7	3.6	382	2 S03762	nucleocapsid prote
24	7	3.6	382	2 E36607	nucleocapsid prote
25	7	3.6	382	2 S47428	nucleocapsid prote
26	7	3.6	405	2 F75596	hypothetical prote
27	7	3.6	407	2 B70185	probable hydroxyme
28	7	3.6	480	2 T20382	hypothetical prote
29	7	3.6	484	2 S67625	probable membrane

hypothetical prote	30	7	3.6	497	2 T29814
PDZ domain family	31	7	3.6	561	2 A87627
probable membrane	32	7	3.6	577	2 S64250
hypothetical prote	33	7	3.6	599	2 T16774
hypothetical prote	34	7	3.6	628	2 B86789
phosphotransferase	35	7	3.6	632	2 F84996
oligopeptidase B (	36	7	3.6	729	2 A97747
diacylglycerol kin	37	7	3.6	788	1 I59282
100K protein - fow	38	7	3.6	791	2 A53691
hypothetical prote	39	7	3.6	798	2 JN0878
conserved hypotet	40	7	3.6	819	2 C84615
hypothetical prote	41	7	3.6	833	2 A90575
leucine-tRNA ligas	42	7	3.6	840	2 T01357
hypothetical prote	43	7	3.6	859	2 F84981
probable serine/th	44	7	3.6	895	2 T32374
DNA-directed DNA p	45	7	3.6	911	2 T01353
probable RND efflu	46	7	3.6	1017	2 T30195
insulin receptor s	47	7	3.6	1025	2 A83186
insulin receptor s	48	7	3.6	1231	2 J30185
insulin receptor s	49	7	3.6	1240	2 JC5209
internal virion pr	50	7	3.6	1242	2 JS0670
rad3 checkpoint pr	51	7	3.6	1318	1 HIBPD7
variant-specific s	52	7	3.6	2386	2 T39911
sperm-specific pro	53	7	3.6	3006	2 T28625
hypothetical prote	54	6	3.1	35	2 B45316
sex-peptide precu	55	6	3.1	56	2 H82647
hypothetical prote	56	6	3.1	60	2 A56547
ribosomal protein	57	6	3.1	64	2 G70731
hypothetical prote	58	6	3.1	75	2 F69139
hypothetical prote	59	6	3.1	84	2 G90044
flagellar motor sw	60	6	3.1	92	2 S34115
cyay protein limpo	61	6	3.1	102	2 F81377
hypothetical prote	62	6	3.1	103	2 C97755
hypothetical prote	63	6	3.1	107	2 F72059
hypothetical prote	64	6	3.1	107	2 E86565
hypothetical prote	65	6	3.1	107	2 F86048
hypothetical prote	66	6	3.1	107	2 B91202
hypothetical prote	67	6	3.1	109	2 G82660
hypothetical prote	68	6	3.1	111	2 S23601
Ig kappa chain V r	69	6	3.1	112	2 A27887
citr protein - Lac	70	6	3.1	112	2 S53881
hypothetical prote	71	6	3.1	114	2 B87129
conserved hypotet	72	6	3.1	118	2 G90227
hypothetical 13.9K	73	6	3.1	120	2 S47819
hypothetical prote	74	6	3.1	120	2 B91188
hypothetical prote	75	6	3.1	120	2 B86035
hypothetical prote	76	6	3.1	122	2 T15766
hypothetical prote	77	6	3.1	122	2 AE0573
conserved hypotet	78	6	3.1	123	1 HSTR2B
histone H2B - rain	79	6	3.1	125	1 HSLUB1
histone H2B.1 - Af	80	6	3.1	125	1 HSLUB2
histone H2B.2 - Af	81	6	3.1	126	2 JH0362
histone H2B.V - ch	82	6	3.1	126	2 S33220
histone H2B.A - Af	83	6	3.1	126	2 S33221
histone H2B.B - Af	84	6	3.1	126	2 T51446
ribosomal protein	85	6	3.1	126	2 T06861
hypothetical prote	86	6	3.1	127	2 AF1253
hypothetical prote	87	6	3.1	131	2 T01293
30S ribosomal prot	88	6	3.1	132	2 D86885
conserved hypotet	89	6	3.1	144	2 H81570
hypothetical prote	90	6	3.1	148	2 F75091
ribosomal protein	91	6	3.1	149	2 G86521
ribosomal protein	92	6	3.1	149	2 G72101
ribosomal protein	93	6	3.1	150	2 A71554
very hypotet	94	6	3.1	150	2 G81705
hypothetical prote	95	6	3.1	150	2 T38477
transcription regu	96	6	3.1	155	2 B71223
hypothetical prote	97	6	3.1	155	2 C75191
hypothetical prote	98	6	3.1	156	2 T32427
conserved hypotet	99	6	3.1	158	2 D81714
allergen RA14B pre	100	6	3.1	166	2 S59922
hypothetical prote	101	6	3.1	168	2 T21692
disease resistance	102	6	3.1	168	2 T08831

**THIS PAGE BLANK (WP10)**

SQ SEQUENCE 367 AA; 40176 MW; ABC91F55B84536EE CRC64;

Query Match 3.6%; Score 7; DB 10; Length 367;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 FIHFENI 26

|||||

Db 260 FIHFENI 266

RESULT 40

O12298

ID O12298 PRELIMINARY; PRT; 376 AA.

AC O12298;

DT 01-JUL-1997 (TReMBLrel. 04, Created)

DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE Nucleocapsid protein.

GN N.

OS feline coronavirus

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.

OX NCBI\_TaxID=12663;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PECV 79-1683;

RA Herrewegh A.A.P.M., Veldkamp A., de Rooij E., Vennema H.,

de Groot R.J.;

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: Y13921; CAA74230.1; ..

DR InterPro: IPR001218; Corona\_nucleocap.

DR Pfam: PF00937; Corona\_nucleoca; 1.

SQ SEQUENCE 376 AA; 42395 MW; F9A8C22350E1D6FD CRC64;

Query Match

Best Local Similarity 3.6%; Score 7; DB 12; Length 376;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135

|||||

Db 338 RKSRKS 344

Search completed: February 26, 2003, 10:01:10

Job time : 67 secs



```

RESULT 36
Q921K3 ID Q921K3 PRELIMINARY; PRT; 339 AA.
AC Q921K3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2002 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RIKEN cDNA 1210001E11 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012039; AAL17808.1; -.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 2.
DR PROSITE: PS0102; RRM; 2.
SQ SEQUENCE 339 AA; 39029 MW; AE8B650D0448A0D CRC64;

Query Match 3.6%; Score 7; DB 11; Length 339;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
DB 241 RKSRKS 247
|||||

RESULT 37
Q93ED3 ID Q93ED3 PRELIMINARY; PRT; 358 AA.
AC Q93ED3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2002 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IMPA.
GN IMPA.
OS Rhizobium leguminosarum (biovar trifolii).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=386;
RN [1]
RP SEQUENCE FROM N.A.
RA Bladergroen M.R.; Badelt K.; Stronk O.P.; Lugtenberg E.J.J.;
RA Spaijk H.P.;
RT "An avirulence locus of a symbiotic Rhizobium leguminosarum strain is
RT involved in temperature-dependent protein secretion."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF361470; AAL17808.1; -.
SQ SEQUENCE 358 AA; 39039 MW; 9F5C107C58D22395 CRC64;

Query Match 3.6%; Score 7; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 QLRLRF 77
DB 87 QLRLRF 93
|||||

RESULT 38
Q8UJF4 ID Q8UJF4 PRELIMINARY; PRT; 364 AA.
AC Q8UJF4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ABC transporter, membrane spanning protein.
GN ATU5523 OR AGR_PAT_778.

```

```

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OG Plasmid AT.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendinning J., Decherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AF008972; AAL46209.1; -.
DR EMBL: AF007920; AAK90899.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 364 AA; 38498 MW; F719F8667AA3CE93 CRC64;

Query Match 3.6%; Score 7; DB 16; Length 364;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 LRLRGFY 78
DB 117 LRLRGFY 123
|||||

RESULT 39
Q9ZPI8 ID Q9ZPI8 PRELIMINARY; PRT; 367 AA.
AC Q9ZPI8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polygalacturonase.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99145623; PubMed=9990097;
RA Bergey D.R., Orozco-Cardenas M., de Moura D.S., Ryan C.A.;
RT "A wound- and systemin-inducible polygalacturonase in tomato leaves."
RL Proc. Natl. Acad. Sci. U.S.A. 96:1756-1760(1999).
CC -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
(POLYGALACTURONASES).
DR EMBL: AF118567; AAD17250.1; -.
DR InterPro: IPR000743; GH28.
DR Pfam: PF00295; Glyco_hydro_28; 1.
DR PROSITE: PS00502; POLYGALACTURONASE; 1.
KW Cell wall; Glycosidase; Hydrolase.

```

```

RA Meikonyan H., Chang W.C., Shapiro J.P., Mahadevappa M.,
RA Fitzpatrick P.A., Kiefer M.C., Tomei D.L., Umansky S.R.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Chang J.T., Esumi N., Moore K., Li Y., Zhang S., Chew C., Goodman B.,
RA Rattner A., Moody S., Stetten G., Campochiaro P.A., Zack D.J.;
RA "Cloning and characterization of a secreted frizzled-related protein
RT that is expressed by the retinal pigment epithelium.";
RT Hum. Mol. Genet. 0:0-0(1999).
DR EMBL; AF017988; AAB70794.1; -.
DR EMBL; AF117758; AAD25052.1; -.
DR InterPro; IPR000024; Fz_domain.
DR InterPro; IPR001134; Netrin_C.
DR Pfam; PF01392; Fz; 1.
DR Pfam; PF01759; NTR; 1.
DR SMART; SM00063; FRI; 1.
DR PROSITE; PS00038; FZ; 1.
SQ SEQUENCE 317 AA; 35577 MW; 05F1FCCCB614F36A CRC64;

Query Match 3.6%; Score 7; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 159 KRKDKTKR 165
Db 236 KRKDKTKR 242
|||||||

RESULT 33
Q96I27 PRELIMINARY; PRT; 334 AA.
AC Q96I27;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to RIKEN cDNA 1200013F24 gene.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLADDER;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006982; AAH06982.1; -.
DR InterPro; IPR001993; Mitoch_carrier.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWNL1.
SQ SEQUENCE 334 AA; 38677 MW; DF31CC8C953E790A CRC64;

Query Match 3.6%; Score 7; DB 4; Length 334;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
Db 42 RKSRKS 48
|||||||

RESULT 34
Q96FN8 PRELIMINARY; PRT; 337 AA.
AC Q96FN8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Unknown (protein for MGC:16799).
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010619; AAHL0619.1; -.
DR InterPro; IPR005302; MOSC.
DR InterPro; IPR005303; MOSC_N.
DR Pfam; PF03473; MOSC; 1.
DR Pfam; PF03476; MOSC_N; 1.
SQ SEQUENCE 337 AA; 37466 MW; 635967B8FE1C8374 CRC64;

Query Match 3.6%; Score 7; DB 4; Length 337;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 107 YFVLENP 113
Db 317 YFVLENP 323
|||||||

RESULT 35
Q9DBP1 PRELIMINARY; PRT; 339 AA.
AC Q9DBP1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1210001E1IRIK protein.
GN 1210001E1IRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Akizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK004831; BAB23599.1; -.
DR MGD; MGI:1915246; 1210001E1IRik.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
SQ SEQUENCE 339 AA; 38997 MW; BC3C9BDA38144A0D CRC64;

Query Match 3.6%; Score 7; DB 11; Length 339;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
Db 241 RKSRKS 247
|||||||

```

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE 1210001E1IRIK protein.  
 GN 1210001E1IRIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10990;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK012605; BAB28351.1; -.  
 DR MGD; MGI:1915246; 1210001E1IRIK.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rtm; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 SQ SEQUENCE 305 AA; 34515 MW; 95F8D2E3D921E500 CRC64;  
  
 Query Match 3.6%; Score 7; DB 11; Length 305;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 129 RKSRSKS 135  
 |||||  
 Db 207 RKSRSKS 213  
  
 RESULT 30  
 Q9PVZ0 PRELIMINARY; PRT; 314 AA.  
 ID Q9PVZ0  
 AC Q9PVZ0;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE C-Jun protein.  
 GN C-JUN.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20500889; PubMed=11044605;  
 RA Knoechel S., Schuler-Wetz A., Knoechel W.;  
 RT "c-Jun (AP-1) activates BMP-4 transcription in Xenopus embryos.";  
 RL Mech. Dev. 98:29-36(2000).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.  
 DR EMBL; AJ243955; CAB51637.1; -.  
 DR HSP; P05412; 1FOS.

DR InterPro; IPR002112; Leuzip\_Jun.  
 DR InterPro; IPR004827; TF\_BZIP.  
 DR Pfam; PF00170; bzip; 1.  
 DR PRINTS; PR00043; LEUZIPPRJUN.  
 DR SMART; SM00338; BRL2; 1.  
 DR PROSITE; PS00036; BZIP\_BASIC; 1.  
 KW DNA-binding; Nuclear protein.  
 SQ SEQUENCE 314 AA; 34556 MW; 73396B48CAC93B3 CRC64;  
  
 Query Match 3.6%; Score 7; DB 13; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 89 RISRLD 95  
 |||||  
 Db 259 RISRLD 265  
  
 RESULT 31  
 Q9XSC1 PRELIMINARY; PRT; 315 AA.  
 ID Q9XSC1  
 AC Q9XSC1;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Secreted frizzled-related protein 5.  
 GN SFRP5.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chang J.T., Esumi N., Moore K., Li Y., Zhang S., Chew C., Goodman B.,  
 RA Rattner A., Moody S., Stetten G., Campochario P.A., Zack D.J.;  
 RT "Cloning and characterization of a secreted frizzled-related protein  
 RT that is expressed by the retinal pigment epithelium."  
 RL Hum. Mol. Genet. 0:0-0(1999).  
 DR EMBL; AF117757; AAD25051.1; -.  
 DR InterPro; IPR000024; Fz\_domain.  
 DR InterPro; IPR001134; Netrin\_C.  
 DR Pfam; PF01392; Fz; 1.  
 DR Pfam; PF01759; NTR; 1.  
 DR SMART; SM00063; FRI; 1.  
 DR PROSITE; PS00038; FZ; 1.  
 SQ SEQUENCE 315 AA; 35279 MW; 8B6E94951060A976 CRC64;  
  
 Query Match 3.6%; Score 7; DB 6; Length 315;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 159 KRKDTKR 165  
 |||||  
 Db 234 KRKDTKR 240  
  
 RESULT 32  
 O14780 PRELIMINARY; PRT; 317 AA.  
 ID O14780  
 AC O14780;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Secreted apoptosis related protein 3.  
 GN SARP3 OR SFRP5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PANCREAS;

```

RESULT 25
Q8VZ95          PRELIMINARY;      PRT;    256 AA.
AC Q8VZ95;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 28.5 kDa protein.
OS Arabidopsis thaliana (Mouse-ear cress).
GN AT3G60600, T4C21.10.
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY065144; AAL38320.1; -.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.
SQ SEQUENCE 256 AA; 28472 MW; 2D17D79F0F736068 CRC64;

Query Match          3.6%; Score 7; DB 10; Length 256;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTPEM 9
Db 117 KEVTPEM 123
|||||

RESULT 26
Q96QK2          PRELIMINARY;      PRT;    276 AA.
AC Q96QK2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Similar to hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RC TISSUE=PROSTATE;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010357; AAH10357.1; -.
DR InterPro; IPR001993; Mitoch_carrier.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
SQ SEQUENCE 276 AA; 31528 MW; 99834B6C3BAF02A CRC64;

Query Match          3.6%; Score 7; DB 4; Length 276;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRSKS 135
Db 42 RKSRSKS 48
|||||

RESULT 27
Q9BI97          PRELIMINARY;      PRT;    282 AA.
ID Q9BI97
AC Q9BI97
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE F17C11.11 protein.
GN F17C11.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z72507; CAC35886.1; -.
SQ SEQUENCE 282 AA; 33577 MW; 212549C6C67EC35E CRC64;

Query Match          3.6%; Score 7; DB 5; Length 282;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 VFGQRF 54
Db 148 VFGQRF 154
|||||

RESULT 28
Q8R8H2          PRELIMINARY;      PRT;    283 AA.
ID Q8R8H2
AC Q8R8H2
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Antirestriction protein.
GN TTE2027.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013153; AAM25204.1; -.
KW Complete proteome.
SQ SEQUENCE 283 AA; 32618 MW; 62611FC401EDB312 CRC64;

Query Match          3.6%; Score 7; DB 16; Length 283;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 NEKSAFD 45
Db 203 NEKSAFD 209
|||||

RESULT 29
Q9CZH6          PRELIMINARY;      PRT;    305 AA.
ID Q9CZH6
AC Q9CZH6;
DT 01-JUN-2001 (TReMBLrel. 17, Created)

```

RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Weber N., Grueninger D., Schmidheini T., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL035528; CAB36833.1; -;  
 DR EMBL; AL161537; CAB78411.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 194 AA; 22117 MW; A50E62552E730F99 CRC64;  
  
 Query Match 3.6%; Score 7; DB 10; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 156 TSKKRD 162  
 DB 98 TSKKRD 104  
  
 RESULT 22  
 Q8SW53  
 ID Q8SW53 PRELIMINARY; PRT; 200 AA.  
 AC Q8SW53;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein ECU03\_0570.  
 GN ECU03\_0570.  
 OS Encephalitozoon cuniculi.  
 OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.  
 OX NCBI\_TaxID=6035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=GB-M1;  
 RC Genoscope;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RX MEDLINE=21576510; PubMed=11719806;  
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,  
 RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissenbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 RT Encephalitozoon cuniculi.";  
 RL Nature 414:450-453(2001).  
 DR EMBL; AL590443; CAD26203.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 200 AA; 23510 MW; 9DB2F9A4ED0E032C CRC64;  
  
 Query Match 3.6%; Score 7; DB 5; Length 200;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 129 RKSRSKS 135  
 DB 191 RKSRSKS 197  
  
 RESULT 23  
 Q9M003  
 ID Q9M003 PRELIMINARY; PRT; 250 AA.  
 AC Q9M003;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE Hypothetical 27.7 kDa protein.  
 GN T4C21\_10.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Choisme N., Robert C., Brottier P., Wincker P., Cattolico L.,  
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Rudd S.,  
 RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL162295; CAB82664.1; -;  
 DR InterPro: IPR000535; MSP\_domain.  
 DR Pfam: PF00635; MSP\_domain; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 250 AA; 27705 MW; E8F8DC2F4D685B66 CRC64;  
  
 Query Match 3.6%; Score 7; DB 10; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 3 KEVTPEM 9  
 DB 117 KEVTPEM 123  
  
 RESULT 24  
 Q9NL00  
 ID Q9NL00 PRELIMINARY; PRT; 253 AA.  
 AC Q9NL00;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE HgTPR5b protein (Fragment).  
 GN HGTPR5B.  
 OS Eptatretus burgeri (Inshore hagfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;  
 OC Myxiniidae; Eptatretinae; Eptatretus.  
 OX NCBI\_TaxID=7764;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2019325; PubMed=10754074;  
 RA Ono-Koyanagi K., Suga H., Katoh K., Miyata T.;  
 RT "Protein tyrosine phosphatases from amphioxus, hagfish, and ray;  
 RT divergence of tissue-specific isoform genes in the early evolution of  
 RT vertebrates.";  
 RL J. Mol. Evol. 50:302-311(2000).  
 DR EMBL; AB033577; BAA95184.1; -;  
 DR HSP; P28827; 1RPM.  
 DR InterPro: IPR000387; TYR\_phosphatase.  
 DR InterPro: IPR000242; Tyr\_PP.  
 DR Pfam: PF00102; Y-phosphatase; 1.  
 DR PRINTS; PR00700; PRTYPHPHTASE.  
 DR SMART; SM00194; PTPC; 1.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 KW Hydrolase.  
 FT NON\_TER  
 SQ SEQUENCE 253 AA; 28315 MW; 27845DD6FC154C96 CRC64;  
  
 Query Match 3.6%; Score 7; DB 13; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 137 SSKSQTR 143  
 DB 122 SSKSQTR 128

RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RL aeolicus";  
 RL Nature 392:353-358(1998).  
 DR EMBL: AE000735; AAC07319.1; -;  
 DR InterPro: IPR000063; ThioRed.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 160 AA; 18519 MW; BCC53BFD9D002201 CRC64;

Query Match 3.6%; Score 7; DB 16; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 RLEDYIK 98  
 |||||  
 Db 151 RLEDYIK 157

RESULT 18  
 Q9NZE5 PRELIMINARY; PRT; 173 AA.  
 ID Q9NZE5  
 AC Q9NZE5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE BM-011.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BONE MARROW;  
 RA Zhao M., Song H., Li N., Peng Y., Han Z., Chen Z.;  
 RT "A novel gene expressed in human bone marrow";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF208853; AAF64267.1; -;  
 SQ SEQUENCE 173 AA; 20900 MW; 20E3D2AE177DBCC3 CRC64;

Query Match 3.6%; Score 7; DB 4; Length 173;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135  
 |||||  
 Db 42 RKSRKS 48

RESULT 19  
 Q939N6 PRELIMINARY; PRT; 176 AA.  
 ID Q939N6  
 AC Q939N6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical 19.0 kDa protein.  
 OS Streptococcus gordonii.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Streptococcus.  
 OX NCBI\_TaxID=1302;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-M99;  
 RA Bensing B.A., Sullam P.M.;  
 RT "An accessory sec locus of Streptococcus gordonii is required for  
 export of GspB and for platelet binding";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY028381; AAL13052.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 176 AA; 19006 MW; 16FC686B3B98297D CRC64;

Query Match 3.6%; Score 7; DB 2; Length 176;

Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 GPQFIHF 23  
 |||||  
 Db 33 GPQFIHF 39

RESULT 20  
 O23308 PRELIMINARY; PRT; 194 AA.  
 ID O23308  
 AC O23308;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Hypothetical 22.0 kDa protein.  
 GN AT4G14520.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Stiekema W., Murphy G., Wambutt R., Pohl T., Terryn N.,  
 RA Kreis M., Kavanagh T., Entian K.D., Rieger M., James R.,  
 RA Puigdomenech P., Hatzopoulos P., Obermaier B., Duesterhoft A.,  
 RA Jones J., Palme K., Ansorge W., Deiseny M., Bancroft I., Mewes H.W.,  
 RA Schueller C., Chaiwatzis N.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z97336; CAB10231.1; -;  
 DR EMBL: AL161539; CAB78494.1; -;  
 DR InterPro: IPR003029; S1.  
 DR SMART: SM00316; S1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 194 AA; 22026 MW; 89822853E434AF10 CRC64;

Query Match 3.6%; Score 7; DB 10; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 ERPHKR 129  
 |||||  
 Db 160 ERPHKR 166

RESULT 21  
 Q9SVP3 PRELIMINARY; PRT; 194 AA.  
 ID Q9SVP3  
 AC Q9SVP3;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Hypothetical 22.1 kDa protein.  
 GN F18A5.80 OR AT4G13690.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Weber N., Grueninger D., Schmidheini T., Bancroft I.,  
 RA Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AP003131; BAB42025.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 127 AA; 15401 MW; 4901EFB0676AE64E CRC64;

Query Match          3.6%; Score 7; DB 16; Length 127;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105
Db      |||||
75 EYCNFGC 81

RESULT 14
Q9QF07 PRELIMINARY; PRT; 129 AA.
ID AC Q9QF07;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE GAG polyprotein (fragment).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LN24;
RX MEDLINE=99412391; PubMed=10482626;
RA Morris A., Marsden M., Halcrow K., Hughes E.S., Brettie R.P.,
RA Bell J.E., Simmonds P.;
RT "Mosaic structure of the human immunodeficiency virus type 1 genome
RT infecting lymphoid cells and the brain: evidence for frequent in vivo
RT recombination events in the evolution of regional populations.";
RL J. Virol. 73:8720-8731(1999).
DR EMBL; AF174941; AAF00287.1; -.
DR InterPro; IPR000071; Retrovir_p17.
DR Pfam; PF00540; Gag_p17; 1.
DR PRINTS; PR00234; HIVMATRIX.
KW AIDS; Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14448 MW; C2234C7181D896B3 CRC64;

Query Match          3.6%; Score 7; DB 15; Length 129;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 IKEEQDK 177
Db      |||||
77 IKEEQDK 83

RESULT 15
Q8STE7 PRELIMINARY; PRT; 137 AA.
ID AC Q8STE7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE D2089.1b protein.
GN D2089.1b.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Swinburne J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;

```

```

RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Gardner A.E.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z36948; CAD30436.1; -.
DR EMBL; Z49911; CAD30436.1; JOINED.
DR EMBL; Z49911; CAD30448.1; -.
DR EMBL; Z36948; CAD30448.1; JOINED.
SQ SEQUENCE 137 AA; 15287 MW; 50F63764F259C1C7 CRC64;

Query Match          3.6%; Score 7; DB 5; Length 137;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 IKFDDER 124
Db      |||||
59 IKFDDER 65

RESULT 16
Q9H6C7 PRELIMINARY; PRT; 151 AA.
ID AC Q9H6C7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA: FLJ22390 fis, clone HRC07810.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026043; BAB15333.1; -.
DR InterPro; IPR005302; MOSC.
DR Pfam; PF03473; MOSC; 1.
SQ SEQUENCE 151 AA; 17097 MW; 47B7C93F0ACE74C1 CRC64;

Query Match          3.6%; Score 7; DB 4; Length 151;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 YFVLENP 113
Db      |||||
131 YFVLENP 137

RESULT 17
O67360 PRELIMINARY; PRT; 160 AA.
ID AC O67360;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein AQ_1343.
GN AQ_1343.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,

```

Q93U42;  
 01-DEC-2001 (TrEMBLrel. 19, Created)  
 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 RNA polymerase beta subunit (fragment).  
 RPOB.  
 OS Carsonella ruddii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.  
 OX NCBI\_TaxID=114186;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Thao M.L., Clark M.A., Burckhardt D.H., Moran N.A., Baumann P.;  
 RT "Phylogenetic analysis of vertically transmitted psyllid endosymbionts  
 (Candidatus Carsonella ruddii) based on atpAGD and rpoC; comparisons  
 with 16S-23S rDNA-derived phylogeny.";   
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
 CC BETA' CHAIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
 DR EMBL: AF268062; AAK55952.1; -;  
 DR InterPro: IPR001572; RNA\_pol\_B.  
 DR Pfam: PF00562; RNA\_pol\_B; 1.  
 DR DNA-directed RNA polymerase; Transcription; Transferase.  
 FT NON\_TER 1  
 SQ SEQUENCE 60 AA; 6786 MW; 90BB0074478AFBFD CRC64;  
 Query Match 3.6%; Score 7; DB 2; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 27 VKSDDIE 33  
 Db 13 VKSDDIE 19  
 RESULT 11  
 ID Q93U40; PRELIMINARY; PRT; 60 AA.  
 AC Q93U40;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE RNA polymerase beta subunit (fragment).  
 GN RPOB.  
 OS Carsonella ruddii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.  
 OX NCBI\_TaxID=114186;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Thao M.L., Clark M.A., Burckhardt D.H., Moran N.A., Baumann P.;  
 RT "Phylogenetic analysis of vertically transmitted psyllid endosymbionts  
 (Candidatus Carsonella ruddii) based on atpAGD and rpoC; comparisons  
 with 16S-23S rDNA-derived phylogeny.";   
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
 CC BETA' CHAIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
 DR EMBL: AF268062; AAK55954.1; -;  
 DR InterPro: IPR001572; RNA\_pol\_B.  
 DR Pfam: PF00562; RNA\_pol\_B; 1.  
 DR DNA-directed RNA polymerase; Transcription; Transferase.  
 KW

FT NON\_TER 1  
 SQ SEQUENCE 60 AA; 6617 MW; 9A34D68E7E563350 CRC64;  
 Query Match 3.6%; Score 7; DB 2; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 27 VKSDDIE 33  
 Db 13 VKSDDIE 19  
 RESULT 12  
 ID Q9YNZ7; PRELIMINARY; PRT; 88 AA.  
 AC Q9YNZ7;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
 DE Hypothetical 9.6 kDa protein.  
 OS Choristoneura fumiferana nuclear polyhedrosis virus (CfMNPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 OC Nucleopolyhedrovirus.  
 OX NCBI\_TaxID=10448;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IRELAND;  
 RA Poloumienko A., Krell P.J.;  
 RT "Identification of the GTA (probable global transactivator) gene of a  
 RT baculovirus pathogenic to the spruce budworm, CfMNPV.";   
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U72240; RAD10319.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 88 AA; 9631 MW; 0D2042034BCC75F5 CRC64;  
 Query Match 3.6%; Score 7; DB 12; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 127 KRRKRS 133  
 Db 18 KRRKRS 24  
 RESULT 13  
 ID Q99VF1; PRELIMINARY; PRT; 127 AA.  
 AC Q99VF1;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein SAV0925.  
 GN SAV0925 OR SA0786.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
 OS Staphylococcus aureus (strain N315).  
 OC Bacteria; Firmicutes; Bacillus/clostridium group; Bacillales;  
 OC Staphylococcus.  
 OX NCBI\_TaxID=158878, 158879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekinizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 RT aureus";   
 RL Lancet 357:1225-1240(2001).  
 DR EMBL: AP003360; BAB57087.1; -;



```

DR EMBL; 275543; CAA99865.1; -.
SQ SEQUENCE 109 AA; 11870 MW; CEI185ACF0F78A47 CRC64;

Query Match
Best Local Similarity 4.2%; Score 8; DB 5; Length 109;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 SKKRKDKT 164
Db 73 SKKRKDKT 80

RESULT 7
Q9CJ28
ID Q9CJ28 PRELIMINARY; PRT; 251 AA.
AC Q9CJ28;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein yb1c.
GN YBIC OR LL0178.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauer S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006255; AAK04276.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 251 AA; 30095 MW; 975D661F0C97E01D CRC64;

Query Match
Best Local Similarity 4.2%; Score 8; DB 16; Length 251;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LNYNKYPG 17
Db 10 LNYNKYPG 17

RESULT 8
Q12418
ID Q12418 PRELIMINARY; PRT; 502 AA.
AC Q12418;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Chromosome XII reading frame ORF YLR094C.
GN GIS3 OR L8004.5 OR YLR094C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Nentwich U., Schwager C., Ansoerge W., Voss H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA Geisel C.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA Taich A., Trevasik E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 273266; CAA97656.1; -.
DR EMBL; U53876; AAB67538.1; -.
DR EMBL; 273267; CAA97658.1; -.
DR SGD; S0004084; GIS3.
SQ SEQUENCE 502 AA; 56356 MW; 62FADCC7FE59337 CRC64;

Query Match
Best Local Similarity 4.2%; Score 8; DB 3; Length 502;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRKS 135
Db 281 RRKRSRKS 288

RESULT 9
Q96J38
ID Q96J38 PRELIMINARY; PRT; 733 AA.
AC Q96J38;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ribosomal protein S6 kinase, 90kD, polypeptide 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002363; AAH02363.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000961; Pkinase_C.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 2.
DR Pfam; PF00433; pkinase_C; 1.
DR ProDom; PD000001; Euk_pkinase; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_2.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN_2.
KW ATP-binding; Kinase; Transferase.
SQ SEQUENCE 733 AA; 83238 MW; 087CFB819A313760 CRC64;

Query Match
Best Local Similarity 4.2%; Score 8; DB 4; Length 733;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRKS 135
Db 19 RRKRSRKS 26

RESULT 10
Q93U42
ID Q93U42 PRELIMINARY; PRT; 60 AA.

```

```

RESULT 3
Q99YU4
ID Q99YU4 PRELIMINARY; PRT; 214 AA.
AC Q99YU4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SPY1534.
GN SPY1534.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006586; AAK34328.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 214 AA; 25536 MW; B054E21499807357 CRC64;

Query Match 6.2%; Score 12; DB 16; Length 214;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 IVGDWGNQRL 74
Db 71 IVGDWGNQRL 82

RESULT 4
Q928N1
ID Q928N1 PRELIMINARY; PRT; 91 AA.
AC Q928N1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein Imo2402.
GN LMO2402 OR LIN2501.
OS Listeria monocytogenes, and
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639, 1642;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=L.monocytogenes, and L.innocua;
RX STRAIN=EGD-E / SEROVAR 1/2A, AND CLIP 11262 / SEROVAR 6A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madsen E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordisiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591983; CAD00480.1; -.
DR EMBL; AL596172; CAC97728.1; -.
DR ListList; LIN02501; -.

DR ListList; LMO02402; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 91 AA; 11001 MW; AEE305DCE28371C CRC64;

Query Match 5.7%; Score 11; DB 16; Length 91;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGCAYFV 109
Db 76 EYCNFGCAYFV 86

RESULT 5
Q9K7D1
ID Q9K7D1 PRELIMINARY; PRT; 93 AA.
AC Q9K7D1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein BH3433.
GN BH3433.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001518; BAB07152.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 93 AA; 11230 MW; IAD87F55CA69E6E1 CRC64;

Query Match 5.7%; Score 11; DB 16; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGCAYFV 109
Db 75 EYCNFGCAYFV 85

RESULT 6
Q21089
ID Q21089 PRELIMINARY; PRT; 109 AA.
AC Q21089;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE K01D12.9 protein.
GN K01D12.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RP Dobson R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RT none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).

```

```

966 6 3..1 982 5 Q960A7
967 6 3..1 982 5 Q9VC93
968 6 3..1 983 5 Q9BI20
969 6 3..1 983 16 Q9HV74
970 6 3..1 985 16 Q92P88
971 6 3..1 1001 5 Q17098
972 6 3..1 1002 3 Q13849
973 6 3..1 1002 10 Q08075
974 6 3..1 1002 10 Q94BS5
975 6 3..1 1004 4 Q9H232
976 6 3..1 1011 5 Q9XVR0
977 6 3..1 1013 3 Q9HF64
978 6 3..1 1020 5 Q9V367
979 6 3..1 1032 16 Q92929
980 6 3..1 1033 5 Q24327
981 6 3..1 1033 5 Q9V643
982 6 3..1 1034 16 Q98IH3
983 6 3..1 1035 5 Q9VJU8
984 6 3..1 1036 16 Q8Y5P1
985 6 3..1 1043 5 Q45232
986 6 3..1 1045 10 Q8RZ01
987 6 3..1 1048 10 Q8W517
988 6 3..1 1050 10 Q81356
989 6 3..1 1052 17 Q8TPC4
990 6 3..1 1053 10 Q9FXK8
991 6 3..1 1054 10 Q9SNY7
992 6 3..1 1057 11 Q91X04
993 6 3..1 1059 3 Q9P8G2
994 6 3..1 1071 16 Q9A0H4
995 6 3..1 1072 11 Q9DBT7
996 6 3..1 1073 16 Q9A4M6
997 6 3..1 1079 10 Q8SLJ6
998 6 3..1 1095 10 Q94J79
999 6 3..1 1095 10 Q943V7
1000 6 3..1 1108 10 Q94ED6

```

## ALIGNMENTS

```

RESULT 1
032127 PRELIMINARY; PRT; 102 AA.
ID Q32127
AC Q32127;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE YUTD protein.
DE YUTD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codoni J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Hega K., Haele J., Harwood C.R., Henaut A.,
RA Halbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

```

```

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipatt A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99120; CAB5221.1; -.
KW Complete proteome.
SQ SEQUENCE 102 AA; 12155 MW; 0551F0BF8AEF3275 CRC64;

Query Match 6.2%; Score 12; DB 16; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGCAFEVL 110
DB 86 EYCNFGCAFEVL 97
|||||
QY 99 EYCNFGCAFEVL 110
DB 86 EYCNFGCAFEVL 97

RESULT 2
Q97RN6 PRELIMINARY; PRT; 170 AA.
ID Q97RN6
AC Q97RN6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SP0767.
DE SP0767.
GN SP0767.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007384; AAK74905.1; -.
DR TIGR; SP0767; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 170 AA; 20318 MW; 80273AB3AD61705E CRC64;

Query Match 6.2%; Score 12; DB 16; Length 170;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 NEQLRLRGFYKD 80
DB 69 NEQLRLRGFYKD 80
|||||
QY 69 NEQLRLRGFYKD 80
DB 69 NEQLRLRGFYKD 80

```

820	6	3.1	680	5	Q9V570	Q9v570 drosophila	893	6	3.1	795	11	088469	088469 rattus norv
821	6	3.1	683	10	Q8V2M4	Q8vzm4 arabidopsis	894	6	3.1	795	11	Q9R1K8	Q9r1k8 rattus norv
822	6	3.1	685	2	Q9F0N9	Q9f0n9 arosiprilla	895	6	3.1	795	11	Q9Z1S3	Q9z1s3 mus musculu
823	6	3.1	685	5	Q961G9	Q961g9 drosophila	896	6	3.1	796	5	Q17323	Q17323 caenorhabdi
824	6	3.1	687	5	Q95R14	Q95r14 drosophila	897	6	3.1	797	4	Q9UNN9	Q9unn9 homo sapien
825	6	3.1	689	3	Q06683	Q06683 saccharomyc	898	6	3.1	797	4	Q95267	Q95267 homo sapien
826	6	3.1	689	4	Q9HB58	Q9hb58 homo sapien	899	6	3.1	803	3	Q74402	Q74402 schizosacch
827	6	3.1	692	6	Q97746	Q97746 canis famil	900	6	3.1	806	11	Q9WVF4	Q9wvf4 mus musculu
828	6	3.1	693	10	Q9S1R0	Q9sir0 arabidopsis	901	6	3.1	809	10	Q8RWQ1	Q8rwq1 arabidopsis
829	6	3.1	695	2	Q9KK51	Q9kk51 streptococc	902	6	3.1	812	4	Q9BW43	Q9bw43 homo sapien
830	6	3.1	695	2	Q9KK36	Q9kk36 streptococc	903	6	3.1	813	16	Q97EC4	Q97ec4 clostridium
831	6	3.1	696	17	Q8TU16	Q8tu16 methanosarc	904	6	3.1	815	10	Q9SCN8	Q9scn8 arabidopsis
832	6	3.1	698	4	Q96MR6	Q96mr6 homo sapien	905	6	3.1	815	16	Q97EX5	Q97ex5 clostridium
833	6	3.1	698	5	Q8T2W9	Q8t2w9 trypanosoma	906	6	3.1	816	4	Q96G51	Q96g51 homo sapien
834	6	3.1	699	2	Q9KK42	Q9kk42 streptococc	907	6	3.1	816	12	Q72163	Q72163 feline astr
835	6	3.1	700	11	Q99K40	Q99k40 mus musculu	908	6	3.1	816	16	Q8XML8	Q8xml8 clostridium
836	6	3.1	702	2	Q93CM6	Q93cm6 bifidobacte	909	6	3.1	817	17	Q9U278	Q9u278 pyrococcus
837	6	3.1	703	5	Q61092	Q61092 trypanosoma	910	6	3.1	820	4	Q60585	Q60585 homo sapien
838	6	3.1	704	4	Q96QI3	Q96qi3 homo sapien	911	6	3.1	823	16	Q9AA14	Q9aa14 caulobacter
839	6	3.1	707	3	Q9P7G0	Q9p7g0 schizosacch	912	6	3.1	824	16	Q8X904	Q8x904 escherichia
840	6	3.1	707	6	Q28297	Q28297 canis famil	913	6	3.1	826	5	Q9XZQ1	Q9xzq1 caenorhabdi
841	6	3.1	707	10	Q94KD3	Q94kd3 arabidopsis	914	6	3.1	826	5	Q9XXH8	Q9xxh8 caenorhabdi
842	6	3.1	708	2	Q82841	Q82841 alteromonas	915	6	3.1	826	5	Q9XTX1	Q9xtx1 caenorhabdi
843	6	3.1	712	16	Q8ZBG2	Q8zbg2 yersinia pe	916	6	3.1	826	16	Q80739	Q80739 arabidopsis
844	6	3.1	713	5	Q19687	Q19687 caenorhabdi	917	6	3.1	826	16	Q92TC2	Q92tc2 rhizobium m
845	6	3.1	714	5	Q45541	Q45541 caenorhabdi	918	6	3.1	827	5	Q96029	Q96029 drosophila
846	6	3.1	717	5	Q8T039	Q8t039 drosophila	919	6	3.1	831	4	Q96MT3	Q96mt3 homo sapien
847	6	3.1	718	5	Q9V594	Q9v594 drosophila	920	6	3.1	832	5	Q9TZK8	Q9tzk8 caenorhabdi
848	6	3.1	720	2	Q9L8J0	Q9l8j0 streptomyc	921	6	3.1	835	13	Q90Z06	Q90z06 xenopus lae
849	6	3.1	720	5	Q9VR99	Q9vr99 drosophila	922	6	3.1	843	10	Q9LZF6	Q9lzf6 arabidopsis
850	6	3.1	722	16	Q926X5	Q926x5 listeria in	923	6	3.1	847	16	Q9RRN5	Q9rrn5 deinococcus
851	6	3.1	724	10	Q9LW21	Q9lw21 arabidopsis	924	6	3.1	850	5	Q95TP1	Q95tp1 drosophila
852	6	3.1	727	17	Q9HHI4	Q9hhi4 halobacteri	925	6	3.1	852	10	Q9ZR08	Q9zr08 arabidopsis
853	6	3.1	728	10	Q9M9S4	Q9m9s4 arabidopsis	926	6	3.1	853	10	Q9FHJ6	Q9fhj6 arabidopsis
854	6	3.1	730	5	Q95QN2	Q95qn2 caenorhabdi	927	6	3.1	857	10	Q9ASK9	Q9ask9 oryza sativ
855	6	3.1	732	5	Q18367	Q18367 caenorhabdi	928	6	3.1	859	3	Q12102	Q12102 saccharomyc
856	6	3.1	733	10	Q9FK08	Q9fkq8 arabidopsis	929	6	3.1	861	10	Q9SDE7	Q9sde7 oryza sativ
857	6	3.1	733	11	Q9D2C0	Q9d2c0 mus musculu	930	6	3.1	862	10	Q43191	Q43191 solanum tub
858	6	3.1	735	13	Q9DD09	Q9ddq9 xenopus lae	931	6	3.1	863	5	Q27267	Q27267 caenorhabdi
859	6	3.1	736	11	Q61164	Q61164 mus musculu	932	6	3.1	865	3	Q8X0B4	Q8x0p4 neurospora
860	6	3.1	737	11	Q9WVF3	Q9wvf3 mus musculu	933	6	3.1	866	13	Q90WV2	Q90wv2 xenopus lae
861	6	3.1	737	11	Q9R1D1	Q9r1d1 rattus norv	934	6	3.1	871	3	Q74653	Q74653 schizosacch
862	6	3.1	739	10	Q9LWY3	Q9lwY3 oryza sativ	935	6	3.1	876	11	Q9WU62	Q9wu62 mus musculu
863	6	3.1	742	5	Q9XZ02	Q9xzq2 caenorhabdi	936	6	3.1	881	16	Q8UJA1	Q8uja1 agrobacteri
864	6	3.1	742	5	Q9XXH9	Q9xxb9 caenorhabdi	937	6	3.1	884	16	Q98E04	Q98e04 rhizobium l
865	6	3.1	743	2	Q8VPJ5	Q8vpj5 clostridium	938	6	3.1	892	5	Q9U299	Q9u299 caenorhabdi
866	6	3.1	743	3	Q9C2D1	Q9c2d1 neurospora	939	6	3.1	893	2	Q93V06	Q93v06 eubacterium
867	6	3.1	743	5	Q9NBV5	Q9nbv5 drosophila	940	6	3.1	897	11	Q70495	Q70495 mus musculu
868	6	3.1	749	5	Q17768	Q17768 caenorhabdi	941	6	3.1	900	10	Q23454	Q23454 arabidopsis
869	6	3.1	752	5	Q21026	Q21026 caenorhabdi	942	6	3.1	901	12	Q8V2F7	Q8v2f7 bean leafro
870	6	3.1	752	5	Q8SUP8	Q8sup8 encephalito	943	6	3.1	904	5	Q45544	Q45544 caenorhabdi
871	6	3.1	752	11	Q55035	Q55035 rattus norv	944	6	3.1	904	5	Q9W0F5	Q9w0t5 drosophila
872	6	3.1	753	5	Q8T2G9	Q8t2g9 dictyosteli	945	6	3.1	911	17	Q975L8	Q975l8 sulfolobus
873	6	3.1	753	16	Q9K3L4	Q9k3l4 streptomyc	946	6	3.1	915	4	Q13827	Q13827 homo sapien
874	6	3.1	754	4	Q13427	Q13427 homo sapien	947	6	3.1	919	4	Q9NQS7	Q9nqs7 homo sapien
875	6	3.1	754	5	Q8SXJ1	Q8sxj1 drosophila	948	6	3.1	923	10	Q49003	Q49003 avena sativ
876	6	3.1	755	6	Q8WP32	Q8wp32 macaca fasc	949	6	3.1	925	10	Q9SIX4	Q9sia4 arabidopsis
877	6	3.1	758	5	Q24086	Q24086 drosophila	950	6	3.1	927	10	Q49004	Q49004 avena sativ
878	6	3.1	759	5	Q9VWZ0	Q9vwz0 drosophila	951	6	3.1	938	5	Q9V3X5	Q9v3x5 drosophila
879	6	3.1	764	4	Q9P2H3	Q9p2h3 homo sapien	952	6	3.1	941	16	Q83382	Q83382 treponema p
880	6	3.1	765	5	Q21478	Q21478 caenorhabdi	953	6	3.1	945	17	Q974N4	Q974n4 tulfofolob
881	6	3.1	768	16	Q8UDA4	Q8uda4 agrobacteri	954	6	3.1	950	13	Q9YHC9	Q9yhc9 xenopus lae
882	6	3.1	773	10	Q22178	Q22178 arabidopsis	955	6	3.1	950	16	Q9HXH0	Q9hxo0 pseudomonas
883	6	3.1	774	3	Q9HEC5	Q9hec5 neurospora	956	6	3.1	958	4	Q8WWI2	Q8wwi2 homo sapien
884	6	3.1	775	12	Q8QY62	Q8qv62 dengue viru	957	6	3.1	959	5	Q9VNI5	Q9vni5 drosophila
885	6	3.1	776	16	Q98A73	Q98a73 rhizobium l	958	6	3.1	960	5	Q9VY25	Q9vy25 drosophila
886	6	3.1	779	16	Q92HZ1	Q92hz1 rickettsia	959	6	3.1	962	5	Q20625	Q20625 caenorhabdi
887	6	3.1	784	16	Q8ZC67	Q8zc67 yersinia pe	960	6	3.1	964	5	Q8T856	Q8t856 dictyosteli
888	6	3.1	787	4	Q9UPY3	Q9upt3 homo sapien	961	6	3.1	967	17	Q8U2E6	Q8u2e6 pyrococcus
889	6	3.1	787	4	Q8TC42	Q8tc42 homo sapien	962	6	3.1	969	17	Q28863	Q28863 archaeoglob
890	6	3.1	787	4	Q8TC27	Q8tc27 homo sapien	963	6	3.1	970	5	Q9VAW1	Q9vaw1 drosophila
891	6	3.1	787	5	Q9VA09	Q9va09 drosophila	964	6	3.1	976	3	Q9C2H8	Q9c2h8 neurospora
892	6	3.1	791	16	Q98QC9	Q98qc9 mycoplasma	965	6	3.1	976	13	Q90ZN9	Q90zn9 brachydanio

674	6	3.1	539	4	Q9HCT8	Q9hct8 homo sapien	747	6	3.1	600	16	Q9HWY8	Q9hwy8 pseudomonas
675	6	3.1	541	10	Q94IJ3	Q94ij3 zea mays (m	748	6	3.1	601	5	O62185	O62185 caenorhabdi
676	6	3.1	542	4	Q9HA87	Q9ha87 homo sapien	749	6	3.1	601	10	Q9M9G3	Q9m9g3 arabidopsis
677	6	3.1	542	4	Q9H9J9	Q9h9j9 homo sapien	750	6	3.1	601	10	Q8VYT2	Q8vvt2 arabidopsis
678	6	3.1	542	16	Q50025	Q50025 mycobacteri	751	6	3.1	606	10	Q9FIW8	Q9fiw8 arabidopsis
679	6	3.1	545	4	Q9NXU7	Q9nxu7 homo sapien	752	6	3.1	607	10	Q9MAQ4	Q9maq4 arabidopsis
680	6	3.1	545	16	Q8WPL8	Q8wpl8 anabaena sp	753	6	3.1	613	10	Q9LQ94	Q9lq94 arabidopsis
681	6	3.1	547	4	Q8WUZ6	Q8wuz6 homo sapien	754	6	3.1	615	10	Q94F62	Q94f62 arabidopsis
682	6	3.1	547	5	Q962D4	Q962d4 ancylostoma	755	6	3.1	617	5	O61395	O61395 caenorhabdi
683	6	3.1	548	2	O54414	O54414 acinetobact	756	6	3.1	617	10	Q9FJP1	Q9fjp1 arabidopsis
684	6	3.1	548	13	Q9PTY5	Q9pty5 xenopus lae	757	6	3.1	618	5	O61394	O61394 caenorhabdi
685	6	3.1	550	4	Q9HCG0	Q9hcg0 homo sapien	758	6	3.1	619	5	Q9WID8	Q9wid8 drosophila
686	6	3.1	551	5	Q95TL9	Q95tl9 drosophila	759	6	3.1	619	11	Q9IYS4	Q9iys4 mus musculu
687	6	3.1	552	4	Q9BR39	Q9br39 homo sapien	760	6	3.1	619	16	O67874	O67874 aquifex aeo
688	6	3.1	552	10	Q9C823	Q9c823 arabidopsis	761	6	3.1	620	8	O99973	O99973 porphyra pu
689	6	3.1	553	10	Q9C823	Q9c823 arabidopsis	762	6	3.1	620	8	O99973	O99973 porphyra pu
690	6	3.1	553	10	O23921	O23921 daucus caro	763	6	3.1	620	10	Q93ZV3	Q93zv3 arabidopsis
691	6	3.1	554	3	O74721	O74721 kluyveromyc	764	6	3.1	621	5	Q9VY09	Q9vy09 drosophila
692	6	3.1	554	5	O16351	O16351 caenorhabdi	765	6	3.1	621	10	Q948Y0	Q948y0 glycine max
693	6	3.1	554	5	O8T434	O8t434 drosophila	766	6	3.1	621	16	O66708	O66708 aquifex aeo
694	6	3.1	556	5	Q9VQH0	Q9vqh0 drosophila	767	6	3.1	622	5	Q95TY0	Q95ty0 drosophila
695	6	3.1	556	10	Q8RYU7	Q8ryu7 oryza sativ	768	6	3.1	622	10	Q93W70	Q93w70 zea mays (m
696	6	3.1	557	16	Q9XOV0	Q9xov0 thermotoga	769	6	3.1	625	10	O94AG2	O94ag2 arabidopsis
697	6	3.1	558	5	Q95R80	Q95r80 thermotoga	770	6	3.1	625	11	O88991	O88991 mus musculu
698	6	3.1	558	5	Q9VCJ5	Q9vcj5 drosophila	771	6	3.1	626	10	Q94IJ5	Q94ij5 zea mays (m
699	6	3.1	558	11	Q91ZR0	Q91zr0 mus musculu	772	6	3.1	626	10	Q94IJ4	Q94ij4 zea mays (m
700	6	3.1	558	17	Q8TWR0	Q8twr0 methanopyru	773	6	3.1	627	11	Q922G0	Q922g0 mus musculu
701	6	3.1	559	10	Q9FZP9	Q9fzp9 glycine max	774	6	3.1	627	4	Q9UK73	Q9uk73 homo sapien
702	6	3.1	559	10	O8S3A6	O8s3a6 lactuca sat	775	6	3.1	627	11	O99NC9	O99nc9 mus musculu
703	6	3.1	560	3	O94667	O94667 schizosacch	776	6	3.1	627	11	O9QZL3	O9qzl3 mus musculu
704	6	3.1	560	10	Q94HQ4	Q94hq4 oryza sativ	777	6	3.1	627	16	O8Z0T5	O8z0t5 salmonella
705	6	3.1	561	5	O8SZS6	O8szs6 drosophila	778	6	3.1	628	10	Q9XIC7	Q9xic7 arabidopsis
706	6	3.1	561	10	Q9C7A9	Q9c7a9 arabidopsis	779	6	3.1	628	10	Q94F63	Q94f63 arabidopsis
707	6	3.1	562	5	Q9U345	Q9u345 caenorhabdi	780	6	3.1	629	3	O8XQ05	O8xq05 neospora
708	6	3.1	562	5	O61396	O61396 caenorhabdi	781	6	3.1	629	5	O22167	O22167 caenorhabdi
709	6	3.1	563	5	O44136	O44136 caenorhabdi	782	6	3.1	631	16	O8XZJ4	O8xzj4 raistonia s
710	6	3.1	564	5	P91218	P91218 caenorhabdi	783	6	3.1	632	10	Q93ZS4	Q93zs4 arabidopsis
711	6	3.1	566	1	O73937	O73937 acidianus a	784	6	3.1	634	16	O8R715	O8r715 thermoanaer
712	6	3.1	566	16	Q9AKQ7	Q9akq7 rhizobium m	785	6	3.1	637	10	Q94DY8	Q94dy8 oryza sativ
713	6	3.1	566	17	Q9C4L9	Q9c4l9 sulfolobus	786	6	3.1	639	16	O8YF90	O8yf90 brucella me
714	6	3.1	567	16	Q98BV2	Q98bv2 rhizobium l	787	6	3.1	639	17	O8ZT28	O8zt28 pyrobaculum
715	6	3.1	568	5	O9U344	Q9u344 caenorhabdi	788	6	3.1	639	17	O9UXE8	O9uxe8 sulfolobus
716	6	3.1	568	5	O02147	O02147 caenorhabdi	789	6	3.1	640	4	Q96PP9	Q96pp9 homo sapien
717	6	3.1	569	16	O8XVH4	O8xvh4 raistonia s	790	6	3.1	640	5	O9VBR1	O9vbr1 drosophila
718	6	3.1	570	10	Q9FLD0	Q9fld0 arabidopsis	791	6	3.1	640	16	Q9JZ98	Q9jz98 neisseria m
719	6	3.1	570	16	Q9PIC5	Q9pic5 campylobact	792	6	3.1	641	2	O93EP2	O93ep2 photorhabdu
720	6	3.1	571	10	Q8THH1	O8thh1 methanosarc	793	6	3.1	645	17	O8TZP8	O8tzp8 pyrococcus
721	6	3.1	571	10	O9SUW9	O9siw9 arabidopsis	794	6	3.1	647	4	O9Y4D2	O9y4d2 homo sapien
722	6	3.1	571	13	O57604	O57604 gallus gall	795	6	3.1	649	10	Q949N7	Q949n7 arabidopsis
723	6	3.1	572	10	Q9LHB3	Q9lhb3 arabidopsis	796	6	3.1	649	16	O83749	O83749 treponema p
724	6	3.1	573	5	O8T022	O8t022 drosophila	797	6	3.1	650	2	O8RR67	O8rr67 actinobacil
725	6	3.1	573	10	Q9SIB0	Q9sib0 arabidopsis	798	6	3.1	650	5	O19878	O19878 caenorhabdi
726	6	3.1	573	10	Q9C7U0	Q9c7u0 arabidopsis	799	6	3.1	652	12	O85056	O85056 atkinsonell
727	6	3.1	574	10	Q9SZG9	Q9szg9 arabidopsis	800	6	3.1	653	2	O34097	O34097 streptococc
728	6	3.1	577	5	O967G8	O967g8 nippostrong	801	6	3.1	653	3	O12171	O12171 saccharomyc
729	6	3.1	577	10	Q9SW08	Q9sw08 arabidopsis	802	6	3.1	656	17	O8TWM1	O8twm1 methanopyru
730	6	3.1	577	16	O8XTF9	O8xtf9 raistonia s	803	6	3.1	657	11	O9EPG7	O9epg7 mus musculu
731	6	3.1	578	5	Q9W324	Q9w324 drosophila	804	6	3.1	660	5	Q9VW90	Q9vw90 drosophila
732	6	3.1	580	10	Q93ZS9	Q93zs9 arabidopsis	805	6	3.1	660	5	O8T431	O8t431 drosophila
733	6	3.1	580	16	O8YTO0	O8yto0 anabaena sp	806	6	3.1	661	5	O9W245	O9w245 drosophila
734	6	3.1	581	10	Q9SVP9	Q9svp9 arabidopsis	807	6	3.1	661	17	Q96XA4	Q96xa4 sulfolobus
735	6	3.1	585	16	Q9KR47	Q9kr47 vibrio chol	808	6	3.1	662	10	Q9SKB8	Q9skb8 arabidopsis
736	6	3.1	586	2	Q9FDT7	Q9fdt7 microcystis	809	6	3.1	665	5	O9VXT5	O9vxt5 drosophila
737	6	3.1	586	10	Q9LY47	Q9ly47 arabidopsis	810	6	3.1	666	16	O8RXC3	O8rcx3 arabidopsis
738	6	3.1	588	10	O22717	O22717 arabidopsis	811	6	3.1	666	16	O8RC42	O8rc42 thermoanaer
739	6	3.1	589	2	Q9LBP2	Q9lbp2 haemophilus	812	6	3.1	667	11	O9R1D9	O9rid9 mus musculu
740	6	3.1	589	10	Q9FLX5	Q9flx5 arabidopsis	813	6	3.1	668	12	O39328	O39328 feline calli
741	6	3.1	591	10	O8S665	O8s665 oryza sativ	814	6	3.1	668	12	O39329	O39329 feline calli
742	6	3.1	592	10	O9LYW0	O9lyw0 arabidopsis	815	6	3.1	673	5	O23706	O23706 carlinus ma
743	6	3.1	592	10	O04096	O04096 arabidopsis	816	6	3.1	674	5	O960F8	O960f8 drosophila
744	6	3.1	597	5	O17562	O17562 caenorhabdi	817	6	3.1	677	3	O74757	O74757 schizosacch
745	6	3.1	598	2	O9A141	O9a141 burkholderi	818	6	3.1	677	17	O8TQS7	O8tqs7 methanosarc
746	6	3.1	598	2	Q93UJ5	Q93uj5 burkholderi	819	6	3.1	678	2	Q93NP3	Q93np3 enterococcu
			599	10	Q9FU99	Q9fu99 arabidopsis							

528	6	3.1	410	16	Q9WA8	Q9ya8 thermotoga	601	6	3.1	480	2	Q9LAX3	Q9lax3 streptococc
529	6	3.1	411	4	Q9BT08	Q9bt08 homo sapien	602	6	3.1	481	2	Q9LAX5	Q9lax5 streptococc
530	6	3.1	411	5	Q22718	Q22718 caenorhabdi	603	6	3.1	482	9	Q8SD07	Q8sd07 pseudomonas
531	6	3.1	411	10	Q949C6	Q949c6 oryza sativ	604	6	3.1	485	10	Q9ZQ07	Q9zq07 arabidopsis
532	6	3.1	412	13	Q9PUB0	Q9pub0 cuturnix co	605	6	3.1	485	12	Q84435	Q84435 paramecium
533	6	3.1	413	5	Q9VJPL	Q9vjpl drosophila	606	6	3.1	487	16	Q69924	Q69924 streptomyce
534	6	3.1	413	11	Q35141	Q35141 rattus norv	607	6	3.1	490	17	Q97216	Q97216 sulfolobus
535	6	3.1	413	11	Q70407	Q70407 mus musculu	608	6	3.1	491	10	Q94455	Q94455 arabidopsis
536	6	3.1	413	13	Q9PUB1	Q9pub1 anolis caro	609	6	3.1	492	5	Q18989	Q18989 caenorhabdi
537	6	3.1	416	16	Q8YRVO	Q8yrvo anabaena sp	610	6	3.1	492	17	Q97Y79	Q97y79 sulfolobus
538	6	3.1	417	4	Q96N00	Q96n00 homo sapien	611	6	3.1	493	16	Q8YWP8	Q8ypw8 anabaena sp
539	6	3.1	418	4	Q8WMM0	Q8wmw0 homo sapien	612	6	3.1	494	2	Q9WXE7	Q9wx7 prevotella
540	6	3.1	418	10	Q9LJ18	Q9lj18 arabidopsis	613	6	3.1	494	11	Q9JK17	Q9jk17 rattus norv
541	6	3.1	418	17	Q8TW51	Q8tw51 methanosarc	614	6	3.1	495	12	Q66458	Q66458 dengue viru
542	6	3.1	421	5	Q9WH79	Q9wh79 drosophila	615	6	3.1	495	12	Q66459	Q66459 dengue viru
543	6	3.1	422	10	Q9LU40	Q9lu40 arabidopsis	616	6	3.1	495	12	Q66460	Q66460 dengue viru
544	6	3.1	423	10	Q9FY60	Q9fy60 arabidopsis	617	6	3.1	495	12	Q66461	Q66461 dengue viru
545	6	3.1	423	10	Q94CG3	Q94cg3 arabidopsis	618	6	3.1	495	12	Q66462	Q66462 dengue viru
546	6	3.1	424	10	Q64420	Q64420 cyanidium c	619	6	3.1	495	12	Q66463	Q66463 dengue viru
547	6	3.1	424	10	Q64419	Q64419 cyanidium c	620	6	3.1	495	12	Q66464	Q66464 dengue viru
548	6	3.1	425	8	Q23890	Q23890 dictyosteli	621	6	3.1	495	12	Q9QCX0	Q9qcx0 dengue viru
549	6	3.1	425	11	Q9CQ38	Q9cq38 mus musculu	622	6	3.1	495	12	Q9QCW6	Q9qcw6 dengue viru
550	6	3.1	428	2	Q9ANY6	Q9any6 enterococcu	623	6	3.1	495	12	Q9QCW5	Q9qcw5 dengue viru
551	6	3.1	430	5	Q95VA6	Q95va6 caenorhabdi	624	6	3.1	495	12	Q9QCW4	Q9qcw4 dengue viru
552	6	3.1	431	10	Q94E25	Q94e25 oryza sativ	625	6	3.1	495	12	Q9QCW2	Q9qcw2 dengue viru
553	6	3.1	432	5	Q9VUX4	Q9vux4 drosophila	626	6	3.1	495	12	Q90203	Q90203 dengue viru
554	6	3.1	434	6	Q9GKV7	Q9gkv7 macaca fasc	627	6	3.1	496	2	Q52249	Q52249 enterococcu
555	6	3.1	435	10	Q9FYM0	Q9fym0 arabidopsis	628	6	3.1	496	2	Q93V07	Q93v07 enterococcu
556	6	3.1	435	10	Q94Y10	Q94y10 arabidopsis	629	6	3.1	496	10	Q9SS55	Q9ss55 arabidopsis
557	6	3.1	437	17	Q8TLL2	Q8tll2 methanosarc	630	6	3.1	497	16	Q9X293	Q9x293 thermotoga
558	6	3.1	438	10	Q9M050	Q9m050 arabidopsis	631	6	3.1	498	16	Q9K994	Q9k994 bacillus ha
559	6	3.1	439	11	Q91W08	Q91w08 mus musculu	632	6	3.1	501	5	Q20179	Q20179 caenorhabdi
560	6	3.1	443	4	Q96F64	Q96f64 homo sapien	633	6	3.1	502	5	Q20280	Q20280 caenorhabdi
561	6	3.1	443	16	Q97NQ5	Q97nq5 streptococc	634	6	3.1	502	5	Q8WQB8	Q8wqb8 caenorhabdi
562	6	3.1	444	10	Q9C8C3	Q9c8c3 arabidopsis	635	6	3.1	503	2	Q93TT3	Q93tt3 vibrio chol
563	6	3.1	444	10	Q9C8C2	Q9c8c2 arabidopsis	636	6	3.1	503	4	Q8TAJ1	Q8taj1 homo sapien
564	6	3.1	445	2	Q68574	Q68574 streptococc	637	6	3.1	504	5	Q16352	Q16352 caenorhabdi
565	6	3.1	448	5	Q9NKE0	Q9nke0 drosophila	638	6	3.1	504	8	Q95GT7	Q95gt7 nepenthes d
566	6	3.1	449	17	Q8TP09	Q8tp09 methanosarc	639	6	3.1	506	17	Q8TMK4	Q8tmk4 methanosarc
567	6	3.1	450	2	Q93H56	Q93h56 streptomyce	640	6	3.1	508	4	Q8WXA9	Q8wxa9 homo sapien
568	6	3.1	451	2	Q9LAB9	Q9lab9 pseudomonas	641	6	3.1	509	10	Q8VZR6	Q8vzr6 arabidopsis
569	6	3.1	451	5	Q20648	Q20648 caenorhabdi	642	6	3.1	510	12	Q91GE8	Q91ge8 epiphyas po
570	6	3.1	452	17	Q97V68	Q97v68 sulfolobus	643	6	3.1	512	3	Q94529	Q94529 schizosacch
571	6	3.1	453	5	Q8SZA0	Q8sza0 drosophila	644	6	3.1	512	10	Q9SV85	Q9sv85 arabidopsis
572	6	3.1	454	16	Q97SF8	Q97sf8 streptococc	645	6	3.1	514	5	Q8SUD0	Q8sud0 encephalito
573	6	3.1	456	11	Q9JLL9	Q9jll9 mus musculu	646	6	3.1	515	5	Q8T8Z5	Q8t8z5 drosophila
574	6	3.1	457	3	Q43087	Q43087 schizosacch	647	6	3.1	516	3	Q96VU3	Q96vu3 lentinula e
575	6	3.1	457	5	Q02096	Q02096 caenorhabdi	648	6	3.1	519	10	Q41714	Q41714 welwitschia
576	6	3.1	457	16	Q92VC3	Q92vc3 rhizobium m	649	6	3.1	520	10	Q9SKG5	Q9skg5 arabidopsis
577	6	3.1	460	10	Q9C8H5	Q9c8h5 arabidopsis	650	6	3.1	520	10	Q9SRL6	Q9srl6 arabidopsis
578	6	3.1	461	2	Q9LAX6	Q9lax6 streptococc	651	6	3.1	520	10	Q9SUD2	Q9sud2 arabidopsis
579	6	3.1	461	16	Q8RB46	Q8rb46 thermoanaer	652	6	3.1	521	10	Q22848	Q22848 arabidopsis
580	6	3.1	463	5	Q9VXJ8	Q9vxj8 drosophila	653	6	3.1	522	5	Q9VN24	Q9vn24 drosophila
581	6	3.1	464	10	Q8S3B7	Q8s3b7 glycine max	654	6	3.1	522	5	Q8T9G3	Q8t9g3 drosophila
582	6	3.1	464	16	Q8YQG4	Q8yqg4 anabaena sp	655	6	3.1	523	3	Q9Y723	Q9y723 irpep lacte
583	6	3.1	465	10	Q8S3B8	Q8s3b8 glycine max	656	6	3.1	523	10	Q9SZC0	Q9szc0 arabidopsis
584	6	3.1	465	16	Q935I8	Q935i8 salmonella	657	6	3.1	524	10	Q9SKG4	Q9skg4 arabidopsis
585	6	3.1	466	3	Q9USN2	Q9usn2 schizosacch	658	6	3.1	525	3	Q12621	Q12621 humicola gr
586	6	3.1	466	4	Q9H9U2	Q9h9u2 homo sapien	659	6	3.1	525	16	Q8RG92	Q8rg92 fusobacteri
587	6	3.1	468	2	Q9LSJ4	Q9lsj4 salmonella	660	6	3.1	527	10	Q9LY45	Q9ly45 arabidopsis
588	6	3.1	469	16	Q9RM32	Q9rm32 vibrio chol	661	6	3.1	527	16	Q98NA7	Q98na7 rhizobium l
589	6	3.1	470	2	Q8VU05	Q8vu05 streptococc	662	6	3.1	528	4	Q9HKU3	Q9hku3 caenorhabdi
590	6	3.1	470	2	Q9AHB1	Q9ahb1 streptococc	663	6	3.1	530	5	Q9BKR8	Q9bkr8 caenorhabdi
591	6	3.1	470	2	Q8RLP8	Q8rlp8 streptococc	664	6	3.1	531	5	Q8T8A7	Q8t8a7 clona intes
592	6	3.1	471	12	Q9ELZ8	Q9elz8 cercopitheci	665	6	3.1	531	6	Q9XS55	Q9xs55 oviss aries
593	6	3.1	472	5	Q20101	Q20101 caenorhabdi	666	6	3.1	532	6	Q9XS56	Q9xs56 oviss aries
594	6	3.1	472	5	Q01631	Q01631 caenorhabdi	667	6	3.1	533	4	Q15317	Q15317 homo sapien
595	6	3.1	473	16	Q8XVP1	Q8xvp1 ralstonia s	668	6	3.1	533	10	Q8RZ85	Q8rz85 oryza sativ
596	6	3.1	476	2	Q9ZG91	Q9z91 pseudomonas	669	6	3.1	533	11	Q9D483	Q9d483 mus musculu
597	6	3.1	477	4	Q96JN6	Q96jn6 homo sapien	670	6	3.1	534	4	Q9Y3R6	Q9y3r6 homo sapien
598	6	3.1	477	17	Q8TQW2	Q8tqw2 methanosarc	671	6	3.1	534	4	Q9BU14	Q9bu14 homo sapien
599	6	3.1	478	5	Q9NRK7	Q9nkr7 drosophila	672	6	3.1	536	3	Q9Y895	Q9y895 volvariella
600	6	3.1	479	2	Q9LAX2	Q9lax2 streptococc	673	6	3.1	538	2	Q9RNB6	Q9rnb6 microcystis

382	6	3.1	314	16	Q92EP8	Q92ep8 listeria mo	455	6	3.1	363	2	Q56612	Q56612 vibrio chol
383	6	3.1	315	4	Q8RVH5	Q8rvh5 homo sapien	456	6	3.1	363	16	Q8UG63	Q8ug63 agrobacteri
384	6	3.1	315	10	Q9SW35	Q9sw35 arabisdopsis	457	6	3.1	365	11	Q9UGJ7	Q9ugj7 mus musculu
385	6	3.1	317	16	Q9KNE7	Q9knf7 vibrio chol	458	6	3.1	367	11	Q9D5P8	Q9d5p8 mus musculu
386	6	3.1	317	16	Q97G43	Q97g43 clostridium	459	6	3.1	369	12	Q9DH64	Q9dh64 meleagrid h
387	6	3.1	318	16	Q97MV2	Q97mv2 clostridium	460	6	3.1	372	4	Q60529	Q60529 homo sapien
388	6	3.1	319	8	Q34805	Q34805 nansenula w	461	6	3.1	373	2	Q9F2H7	Q9f2h7 staphylococ
389	6	3.1	319	16	Q34966	Q34966 bacillus su	462	6	3.1	373	10	Q9LT76	Q9lt76 arabisdopsis
390	6	3.1	321	17	Q82TA3	Q82ta3 pyrobaculum	463	6	3.1	374	16	Q9KFW8	Q9kfw8 bacillus ha
391	6	3.1	321	5	Q20711	Q20711 caenorhabdi	464	6	3.1	375	10	Q9SVA0	Q9sva0 arabisdopsis
392	6	3.1	321	10	Q9XFF0	Q9xf0 avena sativ	465	6	3.1	376	8	Q9MSF2	Q9msf2 nepenthes d
393	6	3.1	321	10	Q23662	Q23662 arabisdopsis	466	6	3.1	378	4	Q9UL09	Q9ul09 homo sapien
394	6	3.1	321	16	Q92VR8	Q92vr8 rhizobium m	467	6	3.1	378	10	Q9CA65	Q9ca65 arabisdopsis
395	6	3.1	324	5	Q9VWR0	Q9vwr0 drosophila	468	6	3.1	379	3	Q9CA41	Q9ca41 alternaria
396	6	3.1	325	16	Q92EQ5	Q92eq5 listeria in	469	6	3.1	379	3	Q9C4A0	Q9c4a0 alternaria
397	6	3.1	325	16	Q8Y9V2	Q8y9v2 listeria mo	470	6	3.1	379	17	Q58234	Q58234 pyrococcus
398	6	3.1	325	16	Q8R9V6	Q8r9v6 thermoanaer	471	6	3.1	379	17	Q97X11	Q97x11 sulfolobus
399	6	3.1	328	4	Q96E84	Q96e84 homo sapien	472	6	3.1	379	17	Q8U3V2	Q8u3v2 pyrococcus
400	6	3.1	328	11	Q9CSU9	Q9csu9 mus musculu	473	6	3.1	380	4	Q9H894	Q9h894 homo sapien
401	6	3.1	328	16	Q9WZB6	Q9wze6 thermotoga	474	6	3.1	382	17	Q9UYG2	Q9uyg2 pyrococcus
402	6	3.1	329	16	Q8U758	Q8u758 agrobacteri	475	6	3.1	383	16	Q25795	Q25795 helicobacte
403	6	3.1	331	3	Q00893	Q00893 colleotetric	476	6	3.1	383	16	Q9ZK33	Q9zk33 helicobacte
404	6	3.1	331	9	Q64366	Q64366 bacterioph	477	6	3.1	385	1	Q59569	Q59569 methanobact
405	6	3.1	332	4	Q60530	Q60530 homo sapien	478	6	3.1	385	4	Q9GZS6	Q9gzs6 homo sapien
406	6	3.1	335	16	Q9Z778	Q9z778 chlamydia p	479	6	3.1	385	4	Q9Y306	Q9y306 homo sapien
407	6	3.1	336	2	Q9ZGL4	Q9zgl4 leptospira	480	6	3.1	385	4	Q9BRX2	Q9brx2 mus musculu
408	6	3.1	336	2	Q9S4H0	Q9s4h0 leptospira	481	6	3.1	385	11	Q91U22	Q91uz2 mus musculu
409	6	3.1	336	3	Q8TFK1	Q8tfk1 debaryomyce	482	6	3.1	386	5	Q36099	Q36099 theileria p
410	6	3.1	338	10	Q9M2D6	Q9m2d6 arabisdopsis	483	6	3.1	387	5	Q9V4V4	Q9v4v4 drosophila
411	6	3.1	339	10	Q9FQ77	Q9fq77 zea mays (m	484	6	3.1	388	5	Q9XTC8	Q9xtc8 caenorhabdi
412	6	3.1	341	16	Q9XNL8	Q9xn18 rhizobium l	485	6	3.1	389	5	Q97253	Q97253 plasmodium
413	6	3.1	342	16	Q9XNS7	Q9xn57 clostridium	486	6	3.1	389	16	Q86581	Q86581 streptomyce
414	6	3.1	343	2	Q31098	Q31098 flavobacter	487	6	3.1	391	5	Q9XVL6	Q9xvl6 caenorhabdi
415	6	3.1	343	5	Q18070	Q18070 caenorhabdi	488	6	3.1	391	10	Q9ZSL4	Q9zsl4 arabisdopsis
416	6	3.1	344	11	Q89037	Q89037 rattus norv	489	6	3.1	392	5	Q9VYI8	Q9vyi8 drosophila
417	6	3.1	344	16	Q34829	Q34829 bacillus su	490	6	3.1	393	11	Q8R4U1	Q8r4ul mus musculu
418	6	3.1	345	3	Q74906	Q74906 schizosacch	491	6	3.1	394	12	Q9Q1Z4	Q9qlz4 rotavirus c
419	6	3.1	345	16	Q99UY5	Q99uy5 staphylococ	492	6	3.1	395	10	Q9LUB8	Q9lub8 arabisdopsis
420	6	3.1	346	10	Q9SYB5	Q9syb5 arabisdopsis	493	6	3.1	395	10	Q9FR67	Q9fr67 cardamine p
421	6	3.1	348	2	Q9ANP0	Q9anp0 bradyrhizob	494	6	3.1	395	12	Q8QLI0	Q8qli0 mamestra co
422	6	3.1	348	16	Q92M44	Q92m44 rhizobium m	495	6	3.1	395	16	Q8KU92	Q8ku92 vibrio chol
423	6	3.1	349	16	Q8ZNK7	Q8znk7 salmonella	496	6	3.1	395	16	Q9CMQ8	Q9cmq8 pasteurella
424	6	3.1	349	16	Q8Z594	Q8z594 salmonella	497	6	3.1	396	12	P88950	P88950 kaposi's sa
425	6	3.1	350	2	Q9RNL7	Q9rnl7 zymomonas m	498	6	3.1	397	10	Q9SFD1	Q9sfd1 arabisdopsis
426	6	3.1	350	13	Q93565	Q93565 xenopus lae	499	6	3.1	399	2	Q9ZHA8	Q9zha8 streptococ
427	6	3.1	350	16	Q929G1	Q929g1 listeria in	500	6	3.1	399	5	Q9VHS8	Q9vhs8 drosophila
428	6	3.1	351	5	Q9N7S0	Q9n7s0 leishmania	501	6	3.1	399	5	Q8SXH3	Q8sxh3 drosophila
429	6	3.1	351	16	Q8U7V6	Q8u7v6 agrobacteri	502	6	3.1	399	16	Q915Q3	Q915q3 pseudomonas
430	6	3.1	352	10	Q9SVY7	Q9svy7 arabisdopsis	503	6	3.1	399	16	Q97RU7	Q97ru7 streptococ
431	6	3.1	353	4	Q96OH5	Q96oh5 homo sapien	504	6	3.1	399	16	Q99SY4	Q99sy4 staphylococ
432	6	3.1	353	4	Q8WWR5	Q8wwr5 homo sapien	505	6	3.1	400	10	Q8S3I8	Q8s3i8 triticum ae
433	6	3.1	353	16	Q9RY70	Q9ry70 deinococcus	506	6	3.1	400	16	Q97KY6	Q97ky6 clostridium
434	6	3.1	353	16	Q8Y564	Q8y564 listeria mo	507	6	3.1	401	2	Q9LAZ2	Q9laz2 streptococ
435	6	3.1	353	17	Q976N6	Q976n6 sulfolobus	508	6	3.1	401	10	Q9SFD0	Q9sfd0 arabisdopsis
436	6	3.1	354	4	Q9UGA4	Q9uga4 homo sapien	509	6	3.1	402	4	Q9NVB6	Q9nvb6 homo sapien
437	6	3.1	354	4	Q8TAN5	Q8tan5 homo sapien	510	6	3.1	402	4	Q9H7U5	Q9huu5 homo sapien
438	6	3.1	354	5	Q8WP45	Q8wp45 halocynthia	511	6	3.1	402	5	Q95Q70	Q95q70 caenorhabdi
439	6	3.1	354	10	Q40067	Q40067 hordeum vul	512	6	3.1	403	5	Q8SVA8	Q8sva8 encephalito
440	6	3.1	354	17	Q28281	Q28281 archaeoglob	513	6	3.1	403	17	Q9YFN7	Q9yfn7 aeropyrum p
441	6	3.1	355	4	Q96C71	Q96c71 homo sapien	514	6	3.1	404	10	Q9C8C4	Q9c8c4 arabisdopsis
442	6	3.1	355	5	Q17757	Q17757 caenorhabdi	515	6	3.1	404	16	Q9A3T0	Q9a3t0 caulobacter
443	6	3.1	355	10	Q9VZF4	Q9vzf4 arabisdopsis	516	6	3.1	406	11	Q9JL35	Q9jl35 mus musculu
444	6	3.1	355	13	Q9W6A4	Q9w6a4 squalus aca	517	6	3.1	406	11	Q8VC71	Q8vc71 mus musculu
445	6	3.1	356	9	Q8SCJ7	Q8scj7 bacterioph	518	6	3.1	406	16	Q8X598	Q8x598 escherichia
446	6	3.1	357	5	Q9GRQ3	Q9grq3 leishmania	519	6	3.1	407	4	Q96GL7	Q96gl7 homo sapien
447	6	3.1	358	3	Q13754	Q13754 schizosacch	520	6	3.1	407	10	Q65879	Q65879 pisum sativ
448	6	3.1	359	2	Q9F329	Q9f329 staphylococ	521	6	3.1	407	16	Q9KUV6	Q9kuv6 vibrio chol
449	6	3.1	359	5	Q9W0Z5	Q9w0z5 drosophila	522	6	3.1	408	4	Q14977	Q14977 homo sapien
450	6	3.1	360	1	P96085	P96085 thermoplasma	523	6	3.1	408	10	Q94A52	Q94a52 arabisdopsis
451	6	3.1	360	17	Q9HI56	Q9hi56 thermoplasma	524	6	3.1	409	4	Q9HA07	Q9ha07 homo sapien
452	6	3.1	361	16	Q9CGI4	Q9cgi4 lactococcus	525	6	3.1	409	4	Q9H0R7	Q9ho7 homo sapien
453	6	3.1	362	16	Q9ZCH3	Q9zch3 rickettsia	526	6	3.1	409	4	Q96EQ1	Q96eq1 homo sapien
454	6	3.1	362	16	Q9ZGC4	Q9zgc4 rickettsia	527	6	3.1	409	4	Q8TB45	Q8tb45 homo sapien

236	3.1	227	17	Q82VD4	Q82vd4 pyrobaculum	309	6	3.1	265	4	Q8Wxf4	Q8wxf4 homo sapien
237	6	228	10	Q8S3K1	Q8s3k1 helianthus	310	6	3.1	265	5	Q9GU88	Q9gu88 paramecium
238	3.1	228	10	Q8S3J9	Q8s3j9 helianthus	311	6	3.1	265	11	Q9CVG5	Q9cvg5 mus musculu
239	6	229	17	Q8ZUQ0	Q8zuq0 pyrobaculum	312	6	3.1	265	16	Q26015	Q26015 helicobacte
240	6	230	16	Q9CGH1	Q9cgh1 lactococcus	313	6	3.1	265	16	Q9ZJD8	Q9zjd8 helicobacte
241	6	231	3	Q93985	Q93985 neocallimas	314	6	3.1	265	16	Q8WJ03	Q8wj03 mycobacteri
242	6	233	2	Q9L568	Q9l568 streptococc	315	6	3.1	266	10	Q9LXY9	Q9lxy9 arabidopsis
243	6	234	4	Q9BVI6	Q9bvi6 homo sapien	316	6	3.1	266	16	Q9JXH1	Q9jxh1 neisseria m
244	6	234	5	Q9XYW7	Q9xyw7 entodinium	317	6	3.1	267	16	Q8YF70	Q8yf70 bruceella me
245	6	236	6	Q9L569	Q9l569 streptococc	318	6	3.1	268	5	Q9GU87	Q9gu87 paramecium
246	6	236	16	Q32238	Q32238 bacillus su	319	6	3.1	268	16	Q9ZTU6	Q9ztu6 rhizobium m
247	6	237	17	Q8ZYL9	Q8zy19 pyrobaculum	320	6	3.1	270	17	Q26323	Q26323 methanobact
248	6	238	2	Q9RBP4	Q9rbp4 rhodococcus	321	6	3.1	270	17	Q8VX0	Q8vxo0 anaebaena sp
249	6	238	10	Q8S3K0	Q8s3k0 helianthus	322	6	3.1	272	16	Q9BR42	Q9br42 homo sapien
250	6	239	17	Q9Y0U6	Q9y0u6 pyrococcus	323	6	3.1	274	17	Q8TJV1	Q8tjv1 methanosarc
251	6	239	17	Q8U143	Q8u143 pyrococcus	324	6	3.1	276	2	Q8WVU1	Q8wvu1 staphylococ
252	6	241	16	Q8RGM2	Q8rgm2 fusobacteri	325	6	3.1	276	10	Q93Y10	Q93y10 arabidopsis
253	6	241	17	Q8TM92	Q8tm92 methanosarc	326	6	3.1	277	2	Q937M1	Q937m1 uncultured
254	6	242	11	Q9D6J2	Q9d6j2 mus musculu	327	6	3.1	277	2	Q937L6	Q937l6 uncultured
255	6	242	17	Q27260	Q27260 methanobact	328	6	3.1	277	2	Q932V8	Q932v8 uncultured
256	6	243	2	Q9L567	Q9l567 streptococc	329	6	3.1	277	3	Q13660	Q13660 schizosacch
257	6	243	2	Q9L564	Q9l564 streptococc	330	6	3.1	277	5	Q8SS59	Q8ss59 encephalito
258	6	243	16	Q8Y3M0	Q8y3m0 listeria mo	331	6	3.1	279	16	Q9KCS3	Q9kcs3 bacillus ha
259	6	244	2	Q9L565	Q9l565 streptococc	332	6	3.1	280	2	Q9EVD3	Q9evd3 neisseria s
260	6	244	17	Q59140	Q59140 pyrococcus	333	6	3.1	280	2	Q93EK4	Q93ek4 neisseria m
261	6	245	5	Q8T9Z7	Q8t9z7 plasmidium	334	6	3.1	281	10	Q8S0I5	Q8s0i5 oryza sativ
262	6	246	2	Q45023	Q45023 borrelia bu	335	6	3.1	283	4	Q96S11	Q96s11 homo sapien
263	6	246	2	Q9L5B4	Q9l5b4 streptococc	336	6	3.1	283	6	Q9XS28	Q9xs28 cercopithe
264	6	246	2	Q8VS45	Q8vs45 borrelia bu	337	6	3.1	283	10	Q9SW52	Q9sw52 arabidopsis
265	6	246	2	Q8VS43	Q8vs43 borrelia bu	338	6	3.1	283	16	Q8ZQS9	Q8zqs9 salmonella
266	6	246	2	Q68233	Q68233 borrelia bu	339	6	3.1	284	5	Q16724	Q16724 caenorhabdi
267	6	247	2	Q9L566	Q9l566 streptococc	340	6	3.1	285	10	Q9FLE7	Q9fle7 arabidopsis
268	6	247	4	Q15262	Q15262 homo sapien	341	6	3.1	286	2	Q87477	Q87477 chlamydia t
269	6	247	16	Q8RAA1	Q8raa1 thermoanaer	342	6	3.1	286	5	Q94390	Q94390 caenorhabdi
270	6	248	4	Q14976	Q14976 homo sapien	343	6	3.1	286	17	Q9Y9J3	Q9y9j3 aeropyrum p
271	6	248	10	Q9XI50	Q9xi50 arabidopsis	344	6	3.1	286	17	Q8TRC9	Q8trc9 methanosarc
272	6	248	16	Q98AL3	Q98al3 rhizobium l	345	6	3.1	287	2	Q9AIX3	Q9aix3 western x p
273	6	249	2	Q9S0B3	Q9s0b3 borrelia bu	346	6	3.1	288	10	Q82396	Q82396 arabidopsis
274	6	249	2	Q68238	Q68238 borrelia bu	347	6	3.1	291	4	Q9BZ75	Q9bz75 homo sapien
275	6	249	2	Q9FBA7	Q9fba7 borrelia he	348	6	3.1	291	4	Q96SB6	Q96sb6 homo sapien
276	6	249	2	Q9L575	Q9l575 streptococc	349	6	3.1	291	4	Q96EM3	Q96em3 homo sapien
277	6	249	2	Q9L570	Q9l570 streptococc	350	6	3.1	292	17	Q27819	Q27819 methanobact
278	6	249	2	Q8VS44	Q8vs44 borrelia bu	351	6	3.1	294	10	Q9S826	Q9s826 arabidopsis
279	6	251	5	Q45595	Q45595 caenorhabdi	352	6	3.1	296	5	Q17357	Q17357 caenorhabdi
280	6	252	16	Q50715	Q50715 borrelia bu	353	6	3.1	296	16	Q8YDX5	Q8ydx5 bruceella me
281	6	252	17	Q9ZV7	Q9zv7 sulfolobus	354	6	3.1	297	2	Q8RWE1	Q8rwe1 bacillus th
282	6	252	17	Q973S4	Q973s4 sulfolobus	355	6	3.1	297	12	Q9LRN3	Q9lrn3 rabies viru
283	6	252	17	Q96Y77	Q96y77 sulfolobus	356	6	3.1	297	12	Q91RL6	Q91rl6 rabies viru
284	6	253	2	Q9ZGC1	Q9zgc1 streptomyce	357	6	3.1	297	12	Q91RG3	Q91rg3 rabies viru
285	6	253	5	Q9BII4	Q9bli4 paragonimus	358	6	3.1	298	4	Q9BRM4	Q9brm4 homo sapien
286	6	254	2	Q9L563	Q9l563 streptococc	359	6	3.1	298	16	Q50682	Q50682 borrelia bu
287	6	255	12	Q89459	Q89459 heliothis a	360	6	3.1	298	16	Q8RBG7	Q8rbg7 thermoanaer
288	6	256	2	Q9L595	Q9l595 streptococc	361	6	3.1	299	5	P91959	P91959 psammecinu
289	6	256	11	Q99J75	Q99j75 mus musculu	362	6	3.1	299	16	Q98H90	Q98h90 rhizobium l
290	6	257	12	Q9E350	Q9e350 maize necro	363	6	3.1	299	17	Q9HQP8	Q9hqp8 halobacteri
291	6	257	13	Q9PD91	Q9pd91 brachydanio	364	6	3.1	301	16	Q8RBH7	Q8rbh7 thermoanaer
292	6	258	5	Q16317	Q16317 caenorhabdi	365	6	3.1	302	16	Q8UKB7	Q8ukb7 agrobacteri
293	6	259	4	Q9NFI7	Q9nfi7 homo sapien	366	6	3.1	303	2	Q67986	Q67986 rhodococcus
294	6	261	4	Q96PI7	Q96pi7 homo sapien	367	6	3.1	303	5	Q18306	Q18306 caenorhabdi
295	6	261	4	Q8WXF0	Q8wxf0 homo sapien	368	6	3.1	303	5	O45673	O45673 caenorhabdi
296	6	261	4	Q8WW25	Q8ww25 homo sapien	369	6	3.1	304	5	Q23337	Q23337 caenorhabdi
297	6	261	16	Q9CJ60	Q9cj60 lactococcus	370	6	3.1	304	10	Q8S3F3	Q8s3f3 arabidopsis
298	6	262	4	Q75494	Q75494 homo sapien	371	6	3.1	305	10	Q93X11	Q93x11 panicum max
299	6	262	5	Q9VBK4	Q9vbk4 drosophila	372	6	3.1	305	17	Q9YDG7	Q9ydg7 aeropyrum p
300	6	262	11	Q9R000	Q9r0u0 mus musculu	373	6	3.1	306	16	Q9AA23	Q9aa23 caulobacter
301	6	262	11	Q88468	Q88468 mus musculu	374	6	3.1	307	10	Q9M2U1	Q9m2u1 arabidopsis
302	6	262	16	Q9JWG4	Q9jwg4 neisseria m	375	6	3.1	309	16	P73682	P73682 synecocyst
303	6	262	16	Q97HG9	Q97hg9 clostridium	376	6	3.1	310	17	Q9UXJ4	Q9uxj4 sulfolobus
304	6	263	12	Q86274	Q86274 avian rotav	377	6	3.1	312	10	Q9X176	Q9x176 arabidopsis
305	6	263	17	Q97B90	Q97b90 thermoplas	378	6	3.1	312	17	Q97A20	Q97a20 thermoplas
306	6	264	10	Q9XG04	Q9xgu4 arabidopsis	379	6	3.1	313	17	Q96YB6	Q96yb6 sulfolobus
307	6	265	2	Q87252	Q87252 lactococcus	380	6	3.1	314	10	Q9LST2	Q9lst2 arabidopsis
308	6	265	2	Q53186	Q53186 rhizobium s	381	6	3.1	314	11	Q9WU66	Q9wu66 mus musculu



90	6	3.1	55	16	Q8XM48	Q8xm48 clostridium	163	6	3.1	153	5	Q8WP53	Q8wp53 solenopsis
91	6	3.1	56	16	Q9PCT3	Q9pct3 xylella fas	164	6	3.1	153	13	Q90XE5	Q90xe5 xenopus lae
92	6	3.1	60	16	Q26313	Q26313 drosophila	165	6	3.1	156	5	Q17325	Q17325 caenorhabdi
93	6	3.1	61	5	Q8STG3	Q8stg3 echinococcu	166	6	3.1	156	5	Q8WRP6	Q8wrp6 solenopsis
94	6	3.1	61	17	Q978C2	Q978c2 thermoplasm	167	6	3.1	158	16	Q9PKX4	Q9pkx4 chlamydia m
95	6	3.1	68	4	Q95048	Q95048 homo sapien	168	6	3.1	160	4	Q9POC0	Q9poc0 homo sapien
96	6	3.1	71	17	Q8TQE0	Q8tqe0 methanosarc	169	6	3.1	161	4	Q9NZA7	Q9nza7 homo sapien
97	6	3.1	74	12	Q39651	Q39651 dengue viru	170	6	3.1	166	4	Q9HA52	Q9ha52 homo sapien
98	6	3.1	74	12	Q39652	Q39652 dengue viru	171	6	3.1	166	10	Q40652	Q40652 oryza sativ
99	6	3.1	76	10	Q9XIK9	Q9xik9 arabidopsis	172	6	3.1	167	5	Q9TZH3	Q9tzh3 caenorhabdi
100	6	3.1	79	5	Q9NE61	Q9ne61 leishmania	173	6	3.1	168	4	Q16043	Q16043 homo sapien
101	6	3.1	82	13	Q8QG36	Q8qgy6 fugu rubrip	174	6	3.1	168	5	Q97194	Q97194 leishmania
102	6	3.1	84	16	Q99RK2	Q99rk2 staphylococ	175	6	3.1	168	5	Q903G9	Q903g9 caenorhabdi
103	6	3.1	90	16	Q9L381	Q9l381 rhizobium m	176	6	3.1	168	10	Q96433	Q96433 glycine max
104	6	3.1	91	16	Q98K21	Q98k21 rhizobium l	177	6	3.1	168	16	Q92110	Q92110 rickettsia
105	6	3.1	95	15	Q91CT6	Q91ct6 chimpanzee	178	6	3.1	169	2	Q8VMP9	Q8vmp9 pseudomonas
106	6	3.1	96	5	Q95QF8	Q95qf8 caenorhabdi	179	6	3.1	170	10	Q8W2E2	Q8w2e2 glycine max
107	6	3.1	96	15	Q90CLO	Q90c10 human immun	180	6	3.1	170	10	Q8W2E1	Q8w2e1 glycine max
108	6	3.1	98	5	Q20560	Q20560 caenorhabdi	181	6	3.1	170	16	Q9JSH0	Q9jsh0 chlamydia p
109	6	3.1	102	16	Q32370	Q32370 campylobact	182	6	3.1	171	10	Q9L2S2	Q9lzs2 arabidopsis
110	6	3.1	104	17	Q96Y32	Q96y32 sulfolobus	183	6	3.1	173	10	Q940P0	Q940p0 arabidopsis
111	6	3.1	105	2	Q9FB68	Q9fb68 lactococcus	184	6	3.1	173	16	Q9PJ08	Q9pj08 campylobact
112	6	3.1	106	2	Q93GG6	Q93gg6 bacillus su	185	6	3.1	174	10	Q9FJR6	Q9fjr6 arabidopsis
113	6	3.1	107	16	Q9Z7V4	Q9z7v4 chlamydia p	186	6	3.1	174	16	Q8FXJ1	Q8fxj1 anabaena sp
114	6	3.1	107	16	Q93GB8	Q93gp8 salmonella	187	6	3.1	175	17	Q97AM6	Q97am6 thermoplasm
115	6	3.1	107	16	Q52124	Q52124 escherichia	188	6	3.1	178	16	Q98LP0	Q98lp0 rhizobium l
116	6	3.1	108	2	Q52854	Q52854 bacillus su	189	6	3.1	180	2	Q9APV2	Q9apv2 pseudomonas
117	6	3.1	108	5	Q8SXF6	Q8sxf6 drosophila	190	6	3.1	180	16	Q8Y6I6	Q8y6i6 listeria mo
118	6	3.1	109	16	Q9PD16	Q9pd16 xylella fas	191	6	3.1	181	10	Q9L2L2	Q9l2l2 arabidopsis
119	6	3.1	111	2	Q53895	Q53895 streptomyce	192	6	3.1	188	6	Q28946	Q28946 sus scrofa
120	6	3.1	111	4	Q96SB5	Q96sb5 homo sapien	193	6	3.1	188	16	Q8RAR5	Q8rar5 thermoanaer
121	6	3.1	111	16	Q8XLB3	Q8xlb3 clostridium	194	6	3.1	189	16	Q8R5R2	Q8r5r2 thermoanaer
122	6	3.1	112	2	Q53512	Q53512 lactococcus	195	6	3.1	189	17	Q980V6	Q980v6 sulfolobus
123	6	3.1	114	16	Q9CBP0	Q9cbp0 mycobacteri	196	6	3.1	191	5	Q95S83	Q95s83 drosophila
124	6	3.1	116	2	Q69105	Q69105 bacillus sp	197	6	3.1	191	5	Q9XTN9	Q9xtn9 meloidogyne
125	6	3.1	117	17	Q973M2	Q973m2 sulfolobus	198	6	3.1	193	4	Q96JQ8	Q96jq8 homo sapien
126	6	3.1	118	5	Q904E5	Q904e5 glossinia pa	199	6	3.1	194	2	Q9L5B5	Q9l5b5 streptococc
127	6	3.1	119	12	Q41527	Q41527 canine herp	200	6	3.1	196	2	Q66087	Q66087 lactococcu
128	6	3.1	120	16	Q8XDH2	Q8xdh2 escherichia	201	6	3.1	196	16	Q97R41	Q97r41 streptococc
129	6	3.1	122	12	Q9WD91	Q9wd91 influenzavi	202	6	3.1	197	5	Q8T671	Q8t671 echinococcu
130	6	3.1	122	16	Q8ZR44	Q8zr44 salmonella	203	6	3.1	198	17	Q59386	Q59386 pyrococcus
131	6	3.1	122	16	Q8Z8M4	Q8z8m4 salmonella	204	6	3.1	201	2	Q45797	Q45797 bacteroides
132	6	3.1	124	5	Q9BL63	Q9bl63 caenorhabdi	205	6	3.1	201	4	Q90UX1	Q90ux1 homo sapien
133	6	3.1	126	5	Q9V8F1	Q9v8f1 drosophila	206	6	3.1	202	11	Q8R0I2	Q8r0i2 mus musculu
134	6	3.1	126	13	Q92130	Q92130 xenopus lae	207	6	3.1	202	16	Q8YNI7	Q8yni7 anabaena sp
135	6	3.1	126	13	Q92131	Q92131 xenopus lae	208	6	3.1	203	16	Q50828	Q50828 borrelia bu
136	6	3.1	127	2	Q9RQH8	Q9rqh8 listeria mo	209	6	3.1	204	5	Q15971	Q15971 drosophila
137	6	3.1	127	9	Q9XIR1	Q9xir1 bacterioph	210	6	3.1	205	2	P71426	P71426 klebsiella
138	6	3.1	127	16	Q8Y771	Q8y771 listeria mo	211	6	3.1	205	2	Q32716	Q32716 klebsiella
139	6	3.1	132	11	Q8RIC2	Q8ric2 mus musculu	212	6	3.1	206	5	Q9VQ06	Q9vq06 drosophila
140	6	3.1	132	16	Q9CDX7	Q9cdx7 lactococcus	213	6	3.1	206	16	Q88045	Q88045 streptomyce
141	6	3.1	133	5	Q8T668	Q8t668 echinococcu	214	6	3.1	207	5	Q45781	Q45781 caenorhabdi
142	6	3.1	133	5	Q8T667	Q8t667 echinococcu	215	6	3.1	209	10	Q88XT3	Q8xrt3 arabidopsis
143	6	3.1	134	5	Q95U06	Q95uq6 branchiost	216	6	3.1	209	16	Q98A43	Q98a43 rhizobium l
144	6	3.1	135	5	Q77007	Q77007 plasmodium	217	6	3.1	211	17	Q97616	Q97616 sulfolobus
145	6	3.1	135	5	Q9TY38	Q9ty38 plasmodium	218	6	3.1	214	2	Q51895	Q51895 prochloroco
146	6	3.1	135	5	Q9TY37	Q9ty37 plasmodium	219	6	3.1	215	13	Q9DFS6	Q9dfs6 brachydanio
147	6	3.1	138	2	Q93SV5	Q93sv5 chlorobium	220	6	3.1	216	8	Q8SL92	Q8sl92 euglena gra
148	6	3.1	141	17	Q8TZG6	Q8tzg6 pyrococcus	221	6	3.1	216	10	Q9C7R9	Q9c7r9 arabidopsis
149	6	3.1	148	17	Q90ZM2	Q90zm2 pyrococcus	222	6	3.1	218	12	Q55589	Q55589 avian rotav
150	6	3.1	150	17	Q972E0	Q972e0 sulfolobus	223	6	3.1	219	5	Q8TIE9	Q8tie9 dictyosteli
151	6	3.1	152	2	Q939G6	Q939g6 pseudomonas	224	6	3.1	220	2	Q9XCW1	Q9xcw1 escherichia
152	6	3.1	153	2	Q48287	Q48287 haemophilus	225	6	3.1	221	11	Q99NX6	Q99nx6 tamias stri
153	6	3.1	153	5	Q8WRQ4	Q8wrq4 solenopsis	226	6	3.1	221	11	Q8VEJ3	Q8vej3 mus musculu
154	6	3.1	153	5	Q8WRQ3	Q8wrq3 solenopsis	227	6	3.1	221	16	Q97KD2	Q97kd2 clostridium
155	6	3.1	153	5	Q8WRQ2	Q8wrq2 solenopsis	228	6	3.1	222	3	Q01599	Q01599 phanerocha
156	6	3.1	153	5	Q8WRQ1	Q8wrq1 solenopsis	229	6	3.1	224	16	Q9PJ95	Q9pj95 campylobact
157	6	3.1	153	5	Q8WRQ0	Q8wrq0 solenopsis	230	6	3.1	225	3	Q74194	Q74194 candida par
158	6	3.1	153	5	Q8WRP9	Q8wrp9 solenopsis	231	6	3.1	225	5	Q967U1	Q967u1 dictyosteli
159	6	3.1	153	5	Q8WRP8	Q8wrp8 solenopsis	232	6	3.1	225	11	Q99NX1	Q99nx1 cricetus
160	6	3.1	153	5	Q8WRP7	Q8wrp7 solenopsis	233	6	3.1	226	5	P91481	P91481 caenorhabdi
161	6	3.1	153	5	Q8WP92	Q8wp92 solenopsis	234	6	3.1	226	10	Q9FY40	Q9fy40 ipomoea bat
162	6	3.1	153	5	Q8WP90	Q8wp90 solenopsis	235	6	3.1	227	17	Q9YB38	Q9yb38 aeropyrum p

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:57:38 : Search time 31 seconds  
(without alignments)  
1276.163 Million cell updates/sec

Title: US-09-868-352-23  
Perfect score: 192  
Sequence: 1 MRKEVTPMLNKNKYPQPF.....EEQDKEMTSKQHLLEVRKN 192

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL\_21.\*

- 1: sp.archaea.\*
- 2: sp.bacteria.\*
- 3: sp.fungi.\*
- 4: sp.human.\*
- 5: sp.invertebrate.\*
- 6: sp.mammal.\*
- 7: sp.mhc.\*
- 8: sp.organelle.\*
- 9: sp.phage.\*
- 10: sp.plant.\*
- 11: sp.rodent.\*
- 12: sp.virus.\*
- 13: sp.vertebrate.\*
- 14: sp.unclassified.\*
- 15: sp.rvirus.\*
- 16: sp.bacteriaph.\*
- 17: sp.archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	6.2	102	16	O32127 bacillus su
2	12	6.2	170	16	Q97RN6 streptococ
3	12	6.2	214	16	Q99YU4 streptococ
4	11	5.7	91	16	Q928N1 listeria mo
5	11	5.7	93	16	Q9K7D1 bacillus ha
6	8	4.2	109	5	Q21089 caenorhabdi
7	8	4.2	251	16	Q9CJ28 lactococcus
8	8	4.2	502	3	Q12418 saccharomyc
9	8	4.2	733	4	Q96J38 homo sapien
10	7	3.6	60	2	Q93U42 carsonella
11	7	3.6	60	2	Q93U40 carsonella
12	7	3.6	88	12	Q9YNZ7 choristoneu
13	7	3.6	127	16	Q99VF1 staphylococ
14	7	3.6	129	15	Q9QFQ7 human immun
15	7	3.6	137	5	Q8STE7 caenorhabdi
16	7	3.6	151	4	Q9H6C7 homo sapien

17	7	3.6	160	16	O67360 aquifex aeo
18	7	3.6	173	4	Q9NZE5 homo sapien
19	7	3.6	176	2	Q939N6 streptococ
20	7	3.6	194	10	O23308 arabidopsis
21	7	3.6	194	10	Q9SVP3 arabidopsis
22	7	3.6	200	5	Q8SW53 encephalito
23	7	3.6	250	10	Q9M003 arabidopsis
24	7	3.6	253	13	Q9NLU0 eptatretus
25	7	3.6	256	10	O8VZ95 arabidopsis
26	7	3.6	276	4	Q960K2 homo sapien
27	7	3.6	282	5	Q9BI97 caenorhabdi
28	7	3.6	283	16	Q8R8H2 thermoanaer
29	7	3.6	305	11	Q9CZH6 mus musculu
30	7	3.6	314	13	Q9PVZ0 xenopus lae
31	7	3.6	315	6	Q9XSC1 bos taurus
32	7	3.6	317	4	O14780 homo sapien
33	7	3.6	334	4	Q961Z7 homo sapien
34	7	3.6	337	4	Q96FN8 homo sapien
35	7	3.6	339	11	Q9DBP1 mus musculu
36	7	3.6	339	11	Q921K3 mus musculu
37	7	3.6	358	2	Q93ED3 rhizobium l
38	7	3.6	364	16	Q8UJF4 agrobacteri
39	7	3.6	367	10	Q9ZPI8 lycopersico
40	7	3.6	376	12	O12298 feline coro
41	7	3.6	382	12	Q88515 transmissib
42	7	3.6	382	12	Q9YRA4 transmissib
43	7	3.6	382	12	O9DY21 transmissib
44	7	3.6	382	12	Q91W03 transmissib
45	7	3.6	395	2	Q9F7Q6 uncultured
46	7	3.6	405	16	Q9RZB8 deinococcus
47	7	3.6	407	16	O51626 borrelia bu
48	7	3.6	445	16	O8XNA1 clostridium
49	7	3.6	452	5	O01159 caenorhabdi
50	7	3.6	484	3	Q12066 saccharomyc
51	7	3.6	497	5	Q18652 caenorhabdi
52	7	3.6	526	3	O74292 gibberella
53	7	3.6	540	17	O971D8 sulfolobus
54	7	3.6	561	16	O9A3Z9 caulobacter
55	7	3.6	626	5	Q22040 caenorhabdi
56	7	3.6	628	16	Q9CFZ7 lactococcus
57	7	3.6	637	17	Q975S8 sulfolobus
58	7	3.6	729	16	Q921P3 rickettsia
59	7	3.6	782	10	Q9FPW0 chlamydomon
60	7	3.6	819	10	O9ZQ47 arabidopsis
61	7	3.6	833	16	Q98Q63 mycoplasma
62	7	3.6	840	4	O14587 homo sapien
63	7	3.6	885	13	Q91615 xenopus lae
64	7	3.6	895	5	O17232 caenorhabdi
65	7	3.6	911	10	O48547 zea mays (m
66	7	3.6	937	11	Q9QVR2 rattus sp.
67	7	3.6	975	5	O44068 stylonychia
68	7	3.6	1003	13	Q9DF49 xenopus lae
69	7	3.6	1017	5	O44055 stylonychia
70	7	3.6	1025	16	Q9HXW4 pseudomonas
71	7	3.6	1111	10	O8S648 oryza sativ
72	7	3.6	1240	13	P79773 gallus gall
73	7	3.6	1251	6	Q28224 cercopithe
74	7	3.6	1308	12	Q9Q9Q6 soil-borne
75	7	3.6	1318	9	Q8W5T9 bacterioph
76	7	3.6	1337	4	Q96RR2 homo sapien
77	7	3.6	1338	4	Q9Y6I5 homo sapien
78	7	3.6	1338	4	Q9BZG0 homo sapien
79	7	3.6	1338	4	Q96RG4 homo sapien
80	7	3.6	1339	4	Q96RG5 homo sapien
81	7	3.6	1472	13	Q90ZAO gallus gall
82	7	3.6	1556	5	O61139 cryptospori
83	7	3.6	1816	12	Q9JAD2 soil-borne
84	7	3.6	3006	5	Q26032 plasmodium
85	6	3.1	27	4	O8WTR6 homo sapien
86	6	3.1	35	2	Q9Z9H9 magnetospir
87	6	3.1	46	6	Q9MZB2 ovis aries
88	6	3.1	48	6	Q9BG7 canis famil
89	6	3.1	53	11	Q922Y6 mus musculu

```

RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kojonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE EARLY ASSEMBLY PROTEINS OF
CC THE 50S RIBOSOMAL SUBUNIT (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L13P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE001610; AAD18400.1; -.
DR EMBL; AE002211; AAF38342.1; -.
DR EMBL; AP002545; BAA98457.1; -.
DR TIGR; CP0515; -.
DR InterPro; IPR001074; Ribosomal_L13.
DR Pfam; PF00572; Ribosomal_L13; 1.
DR ProDom; PD001791; Ribosomal_L13; 1.
DR TIGRFAMs; TIGR01066; rplM_bact; 1.
DR PROSITE; PS00783; RIBOSOMAL_L13; FALSE_NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 149 AA; 16940 MW; 0EFECF02FAA492DF CRC64;

Query Match 3.1%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. NO. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 KRKDTK 164
Db 3 KRKDTK 8

```

Search completed: February 26, 2003, 10:00:20  
Job time : 37 secs

```
KW Hypothetical protein.
SQ SEQUENCE 126 AA; 14349 MW; 51361C2FE7DFB049 CRC64;

Query Match 3.1%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 FVLENP 113
|||||
Db 51 FVLENP 56

RESULT 38
RK12_CYAPA
ID RK12_CYAPA STANDARD; PRT; 126 AA.
AC P48124;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cyanelle 50S ribosomal protein L12.
GN RPL12 OR RPL7.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LB555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LB555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U30821; AAA81204.1; -
CC HSP; P02392; ICTF.
CC InterPro: IPR00206; Ribosomal_L12.
CC Pfam: PF00542; Ribosomal_L12; 1.
CC ProDom: PD001326; Ribosomal_L12; 1.
CC TIGRFAMS; TIGR00855; L12; 1.
CC Ribosomal protein; Cyanelle.
SQ SEQUENCE 126 AA; 13259 MW; D914F5ABE68EF31B CRC64;

Query Match 3.1%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 EKSAFD 45
|||||
Db 59 EKSAFD 64

RESULT 39
RK13_CHLPN
ID RK13_CHLPN STANDARD; PRT; 149 AA.
AC Q92877; Q9JOC2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L13.
GN RPLM OR RL13-OR CPN0247 OR CP0515.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RC MEDLINE=99206606; PubMed=10192388;
RX
```

```
NDK_THEAC
ID NDK_THEAC STANDARD; PRT; 148 AA.
AC Q9HJ59;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
DE (Nucleoside-2-P kinase).
GN NDK OR TALL13.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RC MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES
CC OTHER THAN ATP.
CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate = ADP +
CC nucleoside triphosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE NDK FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL445066; CAC12240.1; -
CC HSP; P15266; INHK.
CC InterPro: IPR001564; NDK.
CC Pfam: PF00334; NDK; 1.
CC PRINTS; PR01243; NUCDPKINASE.
CC ProDom: PD001018; NDK; 1.
CC PROSITE; PS00469; NDP_KINASES; 1.
KW Transferase; Kinase; ATP-binding; Complete proteome.
FT ACT_SITE 116 116 BY SIMILARITY.
SQ SEQUENCE 148 AA; 16657 MW; B09CD0C274BFF9A CRC64;

Query Match 3.1%; Score 6; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 ISRLED 95
|||||
Db 23 ISRLED 28

RESULT 40
RK13_CHLPN
ID RK13_CHLPN STANDARD; PRT; 149 AA.
AC Q92877; Q9JOC2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L13.
GN RPLM OR RL13-OR CPN0247 OR CP0515.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RC MEDLINE=99206606; PubMed=10192388;
RX
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A. (GENE CLUSTER XIH3).  
 RX MEDLINE=86037224; PubMed=3863963;  
 RA Perry M., Thomsen G.H., Roeder R.G.;  
 RT "Genomic organization and nucleotide sequence of two distinct histone  
 gene clusters from *Xenopus laevis*. Identification of novel conserved  
 upstream sequence elements."  
 RL J. Mol. Biol. 185:479-499(1985).  
 RN [2]  
 RP SEQUENCE OF 1-29 AND 62-84.  
 RC TISSUE=Erythrocyte;  
 RX MEDLINE=78144893; PubMed=638193;  
 RA van Helden P., Strickland W.N., Brandt W.F., von Holt C.;  
 RT "Histone H2B variants from the erythrocytes of an amphibian, a  
 reptile and a bird."  
 RL Biochim. Biophys. Acta 533:278-281(1978).  
 CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X03018; CAA26816.1; -;  
 DR EMBL; M21287; AAA49768.1; -;  
 DR PIR; A24510; HSLB1.  
 DR InterPro; IPR000558; Histone\_H2B.  
 DR InterPro; IPR004822; Histone\_core.  
 DR Pfam; PF00125; histone; 1.  
 DR PRINTS; PR00621; HISTONEH2B.  
 DR PRODOM; PD000497; Histone\_H2B; 1.  
 DR SMART; SM00427; H2B; 1.  
 DR PROSITE; PS00357; HISTONE\_H2B; 1.  
 DR Nucleic acid protein; Chromosomal protein; Nucleosome core; DNA-binding;  
 KW Multigene family.  
 FT INIT\_MET 0  
 FT SEQUENCE 125 AA; 13803 MW; DA6C1248381E6F94 CRC64;  
 CC  
 CC Query Match 3.1%; Score 6; DB 1; Length 125;  
 CC Best Local Similarity 100.0%; Pred. No. 50;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 127 KRRKSR 132  
 CC |||||  
 CC Db 28 KRRKSR 33  
 CC  
 CC RESULT 36  
 CC H2B2\_XENLA STANDARD; PRT; 125 AA.  
 CC ID P06900;  
 CC AC P06900;  
 CC DT 01-JAN-1988 (Rel. 06, Created)  
 CC DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 CC DE Histone H2B.2.  
 CC OS *Xenopus laevis* (African clawed frog).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 CC Xenopodinae; Xenopus.  
 CC OX NCBI\_TaxID=8355;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A. (GENE CLUSTER XIH1).  
 CC RX MEDLINE=86037224; PubMed=3863963;

RA Perry M., Thomsen G.H., Roeder R.G.;  
 RT "Genomic organization and nucleotide sequence of two distinct histone  
 gene clusters from *Xenopus laevis*. Identification of novel conserved  
 upstream sequence elements."  
 RL J. Mol. Biol. 185:479-499(1985).  
 CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X03017; CAA26811.1; -;  
 DR EMBL; M21286; AAA49763.1; -;  
 DR PIR; B24510; HSLB2.  
 DR InterPro; IPR000558; Histone\_H2B.  
 DR InterPro; IPR004822; Histone\_core.  
 DR Pfam; PF00125; histone; 1.  
 DR PRINTS; PR00621; HISTONEH2B.  
 DR PRODOM; PD000497; Histone\_H2B; 1.  
 DR SMART; SM00427; H2B; 1.  
 DR PROSITE; PS00357; HISTONE\_H2B; 1.  
 DR Nucleic acid protein; Chromosomal protein; Nucleosome core; DNA-binding;  
 KW Multigene family.  
 FT INIT\_MET 0  
 FT SEQUENCE 125 AA; 13774 MW; 24A035A3F438CA95 CRC64;  
 CC  
 CC Query Match 3.1%; Score 6; DB 1; Length 125;  
 CC Best Local Similarity 100.0%; Pred. No. 50;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 127 KRRKSR 132  
 CC |||||  
 CC Db 28 KRRKSR 33  
 CC  
 CC RESULT 37  
 CC GP48\_BPSP1 STANDARD; PRT; 126 AA.  
 CC ID O48402;  
 CC AC O48402;  
 CC DT 15-DEC-1998 (Rel. 37, Created)  
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 CC DE Putative gene 48 protein.  
 CC GN 48.  
 CC OS Bacteriophage SP01.  
 CC OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 CC SP01-like viruses.  
 CC OX NCBI\_TaxID=10685;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=98327781; PubMed=9657951;  
 RA Stewart C.R., Gaslightwalla I., Hinata K., Krolkowski K.A.,  
 RA Needleman D.S., Peng A.S.-Y., Peterman M.A., Tobias A., Wei P.;  
 RT "Genes and regulatory sites of the 'host-takeover module' in the  
 terminal redundancy of *Bacillus subtilis* bacteriophage SP01."  
 RL Virology 246:329-340(1998).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF031901; AAC29017.1; -;

RT the extremely thermophilic archaeon, Sulfolobus solfataricus P2. ";

RL Genome 43:116-136(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35092 / DSM 1617 / P2;

RX MEDLINE=21332296; PubMed=11427726;

RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,

RA Wayez M.J., Chau-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,

RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,

RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,

RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,

RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,

RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;

RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2. ";

RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).

CC -1- SIMILARITY: BELONGS TO THE UPF0148 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; Y18930; CAB57522.1; -.

DR EMBL; AE006701; AA41078.1; -.

DR InterPro; IPR005355; UPF0148.

DR Pfam; PF03680; UPF0148; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 118 AA; 13445 MW; C92FE8819473D8B2 CRC64;

Query Match 3.1%; Score 6; DB 1; Length 118;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 IVKSDD 31

Db 52 IVKSDD 57

|||||

RESULT 33

YIBI\_ECOLI STANDARD; PRT; 120 AA.

AC P32108;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein yibi.

GN YIBI OR B3598.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=93259920; PubMed=8387990;

RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;

RT "Rhs elements of Escherichia coli K-12: complex composites of shared

RT and unique components that have different evolutionary histories.;"

RL J. Bacteriol. 175:2799-2808(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=94316500; PubMed=8041620;

RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;

RT "Analysis of the Escherichia coli genome. V. DNA sequence of the

RT region from 76.0 to 81.5 minutes.;"

RL Nucleic Acids Res. 22:2576-2586(1994).

CC -1- SIMILARITY: STRONG, TO E.COLI YIAW.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; L19044; AAC95069.1; -.

DR EMBL; U00039; AAB18575.1; -.

DR EMBL; AE000437; AAC76622.1; -.

DR Ecogene; EGI1765; yibi.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 120 AA; 13866 MW; 207110BBE2C03F67 CRC64;

Query Match 3.1%; Score 6; DB 1; Length 120;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 RISRL 94

Db 98 RISRL 103

|||||

RESULT 34

YLB9\_CAEEL STANDARD; PRT; 122 AA.

AC P46583;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Hypothetical 13.8 kDa protein C34E10.9 in chromosome III.

GN C34E10.9

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Kirsten J.;

RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U10402; AAA19071.1; -.

DR Wormpep; C34E10.9; CE01189.

KW Hypothetical protein.

SQ SEQUENCE 122 AA; 13766 MW; F30B8CD153D95598 CRC64;

Query Match 3.1%; Score 6; DB 1; Length 122;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 RQERH 170

Db 46 RQERH 51

|||||

RESULT 35

H2B1\_XENLA STANDARD; PRT; 125 AA.

ID H2B1\_XENLA

AC P02281;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-JAN-1988 (Rel. 06, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Histone H2B.1.

OS Xenopus laevis (African clawed frog).

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE000796; AAB84540.1; -;  
 DR InterPro: IPR000302; KOW\_motif.  
 DR Pfam: PF00467; KOW; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 75 AA; 8335 MW; 4FE84864F6F1258D CRC64;

Query Match 3.1%; Score 6; DB 1; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KSDDIE 33  
 DB 60 KSDDIE 65

## RESULT 30

PHIL\_MYTED  
 ID PHIL\_MYTED STANDARD; PRT; 91 AA.  
 AC Q04621;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE Sperm-specific protein PHI-1.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Sperm;  
 RX MEDLINE=93324352; PubMed=8332480;  
 RA Ruiz-Lara S., Prats E., Casas M.T., Cornudella L.;  
 RT "Molecular cloning and sequence of a cDNA for the sperm-specific  
 RL protein phi 1 from the mussel Mytilus edulis.";  
 RL Nucleic Acids Res. 21:2774-2774(1993).  
 CC -!- FUNCTION: INVOLVED IN NUCLEAR BASIC PROTEIN TRANSITION: HISTONES  
 CC ARE REPLACED BY SPERMATID SPECIFIC PROTEINS WHICH ARE THEMSELVES  
 CC REPLACED BY PROTAMINES IN LATE SPERMATIDS.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: SPERM.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X69718; CAA49375.1; -;  
 DR PIR: S31482; S31482.  
 DR PIR: S34115; S34115.  
 KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 KW Nuclear protein.  
 FT INIT\_MET 0  
 SQ SEQUENCE 91 AA; 10367 MW; FC7E6F5F28344626 CRC64;

Query Match 3.1%; Score 6; DB 1; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 KRSRKS 135  
 DB 9 KRSRKS 14

## RESULT 31

RT "Gene content and organization of a 281-kbp contig from the genome of

CYAY\_RICCN  
 ID CYAY\_RICCN STANDARD; PRT; 103 AA.  
 AC Q92IH7;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cyay protein.  
 GN CYAY OR RC0443.  
 OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsieae; Rickettsia.  
 OX NCBI\_TaxID=781;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Malish 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
 RL Science 293:2093-2098(2001).  
 CC -!- SIMILARITY: BELONGS TO THE FRATAXIN FAMILY

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE008608; AAL02981.1; -;  
 DR InterPro: IPR002908; Frataxin\_like.  
 DR Pfam: PF01491; Frataxin\_Cyay; 1.  
 DR ProDom: PD006646; Frataxin\_like; 1.  
 DR PROSITE: PS01344; FRATAXIN\_1; 1.  
 DR PROSITE: PS0810; FRATAXIN\_2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 103 AA; 11689 MW; 1B3017B3E27FA40D CRC64;

Query Match 3.1%; Score 6; DB 1; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 EEOQKE 178  
 DB 22 EEOQKE 27

## RESULT 32

Y781\_SULSO  
 ID Y781\_SULSO STANDARD; PRT; 118 AA.  
 AC Q9UXG8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein SSO0781.  
 GN SSO0781 OR C40.017.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=20165948; PubMed=10701121;  
 RA Charlebois R.L., Singh R.K., Chan-Weiher C.C.-Y., Allard G., Chow C.,  
 RA Confalonieri F., Curtis B., Duquet M., Erauso G., Faguy D.,  
 RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,  
 RA Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,  
 RA St Jean A., van der Oost J., Young F., Zivanovic Y., Doolittle W.F.,  
 RA Ragan M.A., Sensen C.W.;

```

Query Match      3.6%; Score 7; DB 1; Length 2386;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 33 EFQLVIN 39
    |
Db 1299 EFQLVIN 1305

RESULT 27
PHIL_MYTCA
ID PHIL_MYTCA STANDARD; PRT; 35 AA.
AC P35422;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE Sperm-specific protein PHI-1 (PL-III) (Sperm-specific protamine-like
DE protein) (Fragment).
OS Mytilus californianus (California mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloida; Mytilidae; Mytilus.
OX NCBI_TaxID=6549;
RN [1]
RP SEQUENCE.
RC TISSUE=Sperm; PubMed=7677995;
RX Carlos S., Jutgier L., Borrell I., Hunt D.F., Ausio J.;
RT "Sequence and characterization of a sperm-specific histone H1-like
RT protein of Mytilus californianus.";
RL J. Biol. Chem. 268:185-194(1993).
CC -1- FUNCTION: INVOLVED IN NUCLEAR BASIC PROTEIN TRANSITION: HISTONES
CC ARE REPLACED BY SPERMATID SPECIFIC PROTEINS WHICH ARE THEMSELVES
CC REPLACED BY PROTAMINES IN LATE SPERMATIDS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: SPERM.
DR PIR; B45316; B45316.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Nuclear protein.
ET NON_TER 35 35
SQ SEQUENCE 35 AA; 3759 MW; 97459D27A76F4BD1 CRC64;

Query Match      3.1%; Score 6; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 KSRKS 135
    |
Db 8 KSRKS 13

RESULT 28
YM83_MYCTU
ID YM83_MYCTU STANDARD; PRT; 64 AA.
AC Q50682;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Very hypothetical protein RV2283.
GN RV2283 OR MT2341.1 OR MTCY339.27C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
"Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
Nature 393:537-544(1998).
(2)
SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z77163; CAB00985.1; -;
DR EMBL; AE007077; -; NOT_ANNOTATED_CDS.
DR TIGR; MT2341.1; -;
DR TubercuList; RV2283; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 64 AA; 6557 MW; 9B7C5467FB8451AC CRC64;

Query Match      3.1%; Score 6; DB 1; Length 64;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 ASTRK 86
    |
Db 33 ASTRK 38

RESULT 29
R14E_METH
ID R14E_METH STANDARD; PRT; 75 AA.
AC O26139;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L14E.
GN RPL14E OR MTH31.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT J. Bacteriol. 179:7135-7155(1997).
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- SIMILARITY: BELONGS TO THE L14E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```



```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wallis J.M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PUTATIVE RNA HELICASE INVOLVED IN THE SECOND STEP OF RNA
CC SPlicing (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- DOMAIN: COMPOSED OF TWO SIMILAR DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SKI2 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL110485; CAB60351.1; -.
CC WormPep; Y46G5A.4; CE21971.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR001650; Helicase_C.
CC Pfam: PF00270; DEAD; 2.
CC Pfam: PF00271; helicase_C; 1.
CC Pfam: PF02889; Sec63; 2.
CC SMART; SM00487; DEXDC; 2.
CC SMART; SM00490; HELICC; 1.
CC KW Hypothetical protein; Helicase; mRNA processing; mRNA splicing;
CC KW Spliceosome; Nuclear protein; ATP-binding; Repeat.
FT DOMAIN 497 897
FT NP_BIND 1344 1738
FT ATP (POTENTIAL).
FT NP_BIND 1344 1351
FT ATP (POTENTIAL).
FT SITE 609 612
FT DEIH BOX.
FT SITE 1448 1451
FT DDLH BOX.
FT DOMAIN 221 228
FT POLY-GLU.
SQ SEQUENCE 2145 AA; 243829 MW; 2C1EE0F69583891F CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 LVINEKS 42
Db 2084 LVINEKS 2090
|||||||

RESULT 26
RAD3_SCHPO STANDARD; PRT; 2386 AA.
AC Q02099; O9UUM1; O92391;
DT 1-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA repair protein rad3.
GN RAD3 OR SPBC216.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=97133293; PubMed=8978690;
RA Bentley N.J., Holtzman D.A., Flaggs G., Keegan K.S., DeMaggio A.,
RA Ford J.C., Hoekstra M., Carr A.M.;
RT "The Schizosaccharomyces pombe rad3 checkpoint gene.";
RL EMBO J. 15:6641-6651(1996).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkkaert G., Aert R., Robben J., Grymonprez B.,
RA Welljans I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fricz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux C., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [3]
RP SEQUENCE OF 711-1781 FROM N.A.
RX MEDLINE=93012978; PubMed=1398093;
RA Seaton B.L., Yucel J., Sunnerhagen P., Subramani S.;
RT "Isolation and characterization of the Schizosaccharomyces pombe rad3
RL gene, involved in the DNA damage and DNA synthesis checkpoints.";
CC Gene 119:83-89(1992).
CC -!- FUNCTION: INVOLVED IN G2 ARREST FOLLOWING DNA DAMAGE WHERE IT
CC PHOSPHORYLATES CHK1. IT IS ALSO INVOLVED IN THE DEPENDENCE OF
CC MITOSIS ON THE COMPLETION OF DNA REPLICATION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y09076; CAA70297.1; -.
CC EMBL: U76307; AAC49607.1; -.
CC EMBL: AL049558; CAB40165.1; -.
CC EMBL: X63544; CAA45106.1; -.
CC PIR: S25834; S25834.
CC InterPro: IPR003151; FAT.
CC InterPro: IPR003152; FATC.
CC InterPro: IPR000403; PI3_P14_kinase.
CC InterPro: IPR001440; TPR.
CC Pfam: PF00454; PI3_P14_kinase; 1.
CC Pfam: PF02259; FAT; 1.
CC Pfam: PF02260; FATC; 1.
CC SMART; SM00146; PI3KC; 1.
CC PROSITE; PS00915; PI3_4_KINASE_1; FALSE_NEG.
CC PROSITE; PS00916; PI3_4_KINASE_2; FALSE_NEG.
CC PROSITE; PS0290; PI3_4_KINASE_3; 1.
CC Transferrase; Kinase; DNA damage; DNA synthesis; DNA repair;
KW Nuclear protein.
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=97133293; PubMed=8978690;
RA Bentley N.J., Holtzman D.A., Flaggs G., Keegan K.S., DeMaggio A.,
RA Ford J.C., Hoekstra M., Carr A.M.;
RT "The Schizosaccharomyces pombe rad3 checkpoint gene.";
RL EMBO J. 15:6641-6651(1996).
RN [2]

```

RA Glasheen E., Lane W.S., Pierce J.H., White M.F.;  
 RL "Role of IRS-2 in insulin and cytokine signalling.";  
 CC Nature 377:173-177(1995).  
 CC -1- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES  
 CC BY INSULIN.  
 CC -1- TISSUE SPECIFICITY: SKELETAL MUSCLE, LUNG, BRAIN, LIVER, KIDNEY,  
 CC HEART AND SPLEEN.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PTB DOMAIN.  
 CC HSP; P35568; IIRS.  
 DR MGD; MGI:109334; IIRS2.  
 DR InterPro; IPR002404; Insulin\_receptsl.  
 DR InterPro; IPR001849; PH.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF02174; IRS; 1.  
 DR PRINTS; PR00628; INSULINRSI.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00310; PTBI; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 KW Phosphorylation.  
 FT DOMAIN 16 144 PH.  
 FT DOMAIN 187 300 PTB.  
 FT MOD\_RES 536 536 (BY SIMILARITY).  
 FT MOD\_RES 649 649 (BY SIMILARITY).  
 FT MOD\_RES 671 671 (BY SIMILARITY).  
 FT MOD\_RES 911 911 (BY SIMILARITY).  
 FT MOD\_RES 970 970 (BY SIMILARITY).  
 FT MOD\_RES 1242 1242 (BY SIMILARITY).  
 FT MOD\_RES 1303 1303 (BY SIMILARITY).  
 FT DOMAIN 19 28 POLY-ASN.  
 FT DOMAIN 444 449 POLY-SER.  
 FT DOMAIN 638 641 POLY-SER.  
 FT DOMAIN 936 939 POLY-SER.  
 SQ SEQUENCE 1321 AA; 136526 MW; 5069CE9D614960C7 CRC64;  
 Query Match 3.6%; Score 7; DB 1; Length 1321;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 132 RSKSQSS 138  
 Db 300 RSKSQSS 306  
 RESULT 24  
 IRS2\_HUMAN STANDARD; PRT; 1324 AA.  
 AC Q9Y4H2;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Insulin receptor substrate-2 (IRS-2).  
 GN IRS2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97460123; PubMed=9312143;  
 RA Ogihara T., Isobe T., Ichimura T., Taoka M., Funaki M., Sakoda H.,  
 RA Onishi Y., Inukai K., Anai M., Fukushima Y., Kikuchi M., Yazaki Y.,  
 RA Oka Y., Asano T.;  
 RT "14-3-3 protein binds to insulin receptor substrate-1, one of the  
 RT binding sites of which is in the phosphorytyrosine binding domain.";  
 RL J. Biol. Chem. 272:25267-25274(1997).

CC -1- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES  
 CC BY INSULIN.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PTB DOMAIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AB000732; BAA24500.1; .  
 DR HSSP; P35568; IIRS.  
 DR Genew; HGNC:6126; IRS2.  
 DR MIM; 600797; .  
 DR InterPro; IPR002404; Insulin\_receptsl.  
 DR InterPro; IPR001849; PH.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF02174; IRS; 1.  
 DR PRINTS; PR00628; INSULINRSI.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00310; PTBI; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 KW Phosphorylation.  
 FT DOMAIN 16 144 PH.  
 FT DOMAIN 190 303 PTB.  
 FT MOD\_RES 540 540  
 FT MOD\_RES 653 653 (BY INSR) (BY  
 FT MOD\_RES 675 675 (BY INSR) (BY  
 FT MOD\_RES 919 919 (BY INSR) (BY  
 FT MOD\_RES 978 978 (BY INSR) (BY  
 FT MOD\_RES 1253 1253 (BY INSR) (BY  
 FT DOMAIN 19 28 POLY-ASN.  
 FT DOMAIN 371 380 POLY-ALA.  
 FT DOMAIN 447 452 POLY-SER.  
 FT DOMAIN 460 467 POLY-PRO.  
 FT DOMAIN 533 537 POLY-GLY.  
 FT DOMAIN 642 645 POLY-SER.  
 FT DOMAIN 694 701 POLY-ALA.  
 FT DOMAIN 944 947 POLY-SER.  
 FT DOMAIN 1031 1038 POLY-PRO.  
 FT DOMAIN 1265 1278 POLY-PRO.  
 SQ SEQUENCE 1324 AA; 136482 MW; 3D7B4AB2AE45104 CRC64;  
 Query Match 3.6%; Score 7; DB 1; Length 1324;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 132 RSKSQSS 138  
 Db 303 RSKSQSS 309  
 RESULT 25  
 U520\_CAEEL STANDARD; PRT; 2145 AA.  
 ID U520\_CAEEL  
 AC Q9U2G0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative U5 small nuclear ribonucleoprotein 200 kDa helicase.  
 GN Y46GSA.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.

```
CC EMBL; S85963; AAB21608.1; -
DR EMBL; S62539; AAB27175.1; -
DR PIR; JS0670; JS0670.
DR PDB; 1JRS; 15-MAY-97.
DR Genew; HGNC:6125; IRS1.
DR MIM; 147545; -.
DR InterPro; IPR002404; Insulin_receptsl.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF02174; IRS; 1.
DR PRINTS; PR00628; INSULINRSI.
DR SMART; SM00233; PH; 1.
DR SMART; SM00310; PTBI; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR Phosphorylation; Polymorphism; Diabetes mellitus; Disease mutation;
KW 3D-structure.
FT DOMAIN 12 115 PH.
FT DOMAIN 157 267 PTH.
FT DOMAIN 128 134 POLY-GLY.
FT DOMAIN 391 398 POLY-SER.
FT DOMAIN 680 686 POLY-SER.
FT DOMAIN 807 815 POLY-SER.
FT DOMAIN 877 882 POLY-GLN.
FT DOMAIN 1055 1058 POLY-SER.
FT DOMAIN 1126 1130 POLY-GLY.
FT DOMAIN 1131 1135 POLY-SER.
FT DOMAIN 1197 1207 POLY-PRO.
FT MOD_RES 99 99 PHOSPHORYLATION (BY CK2)
FT MOD_RES 455 455 (BY SIMILARITY).
FT MOD_RES 612 612 PHOSPHORYLATION (BY INSR)
FT MOD_RES 632 632 PHOSPHORYLATION (BY INSR)
FT MOD_RES 896 896 PHOSPHORYLATION (BY INSR)
FT MOD_RES 941 941 PHOSPHORYLATION (BY INSR)
FT MOD_RES 989 989 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1179 1179 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1229 1229 PHOSPHORYLATION (BY INSR)
FT VARIANT 512 512 A -> P.
FT VARIANT 723 723 /FTID=VAR_005299.
FT VARIANT 971 971 MISSING (IN NIDDM).
FT VARIANT 1043 1043 /FTID=VAR_005301.
FT VARIANT 1095 1095 G -> R.
FT VARIANT 1095 1095 /FTID=VAR_005300.
FT VARIANT 1095 1095 S -> Y (IN NIDDM).
FT VARIANT 1095 1095 /FTID=VAR_005302.
FT VARIANT 1095 1095 C -> Y (IN NIDDM).
FT VARIANT 1095 1095 /FTID=VAR_005303.
FT CONFLICT 134 134 G -> GG (IN REF. 2).
FT CONFLICT 362 362 S -> R (IN REF. 2).
FT CONFLICT 384 384 P -> R (IN REF. 2).
SQ SEQUENCE 1242 AA; 131590 MW; 3C0EFD9E32B3E64A CRC64;

Query Match 3.6%; Score 7; DB 1; Length 1242;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 267 RSKSQSS 273
IIIIIIII

RESULT 22
VIVD_BPT7 STANDARD; PRT; 1318 AA.
ID VIVD_BPT7
```

```
AC P03726;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Internal virion protein D.
GN 16.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10760;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83241725; PubMed=6864790;
RT Dunn J.J., Studier F.W.;
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
RT locations of T7 genetic elements.";
RL J. Mol. Biol. 166:477-535(1983).
RN [2]
RP SIMILARITY TO SLT.
RX MEDLINE=94262160; PubMed=8203016;
RA Koonin E.V., Rudd K.E.;
RT "A conserved domain in putative bacterial and bacteriophage
RT transglycosylases.";
RL Trends Biochem. Sci. 19:106-107(1994).
CC -!- FUNCTION: MAY BE INVOLVED IN THE LYSIS OF THE BACTERIAL CELL
CC WALL DURING THE RELEASE OF THE PHAGE PROGENY.
CC -!- SIMILARITY: BELONGS TO THE SLT FAMILY OF TRANSGLYCOSYLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01146; CAA24434.1; -.
DR PIR; A04352; HIBPD7.
DR PIR; S42332; S42332.
DR InterPro; IPR000189; SLT_domain.
DR Pfam; PF01464; SLT; 1.
DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
KW Cell wall; Hydrolase; Glycosidase.
FT DOMAIN 24 111 SLT-TYPE DOMAIN.
FT ACT_SITE 37 37 BY SIMILARITY.
SQ SEQUENCE 1318 AA; 143838 MW; 51A0AAA920CBF210 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 1318;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTPEM 9
IIIIIIII
Db 720 KEVTPEM 726

RESULT 23
IRS2_MOUSE STANDARD; PRT; 1321 AA.
ID IRS2_MOUSE
AC P81122;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Insulin receptor substrate-2 (IRS-2) (4PS).
GN IRS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=95405472; PubMed=7675087;
RA Sun X.J., Wang L.-M., Zhang Y., Yenush L., Myers M.G. Jr.,
```

```

RT cloning.*;
RL Biochim. Biophys. Acta 1172:323-326(1993).
CC -!- FUNCTION: MAY MEDATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC BY INSULIN. WHEN PHOSPHORYLATED BY THE INSULIN RECEPTOR BINDS
CC SPECIFICALLY TO VARIOUS CELLULAR PROTEINS CONTAINING SH2 DOMAINS
CC SUCH AS PHOSPHATIDYLINOSITOL 3-KINASE P85 SUBUNIT OR GRB-2.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L24563; AAA39335.1; -.
DR EMBL; X69722; CAA49378.1; -.
DR PIR; S43514; S43514.
DR HSP; P35568; IIRS.
DR MGD; MGI:99454; Irs1.
DR InterPro; IPR002404; Insulin_receptS1.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF02174; IRS; 1.
DR PRINTS; PR00628; INSULINRS1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00310; PTB; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Phosphorylation.
FT DOMAIN 12 115 PH.
FT DOMAIN 152 262 PTB.
FT DOMAIN 675 680 POLY-SER.
FT DOMAIN 872 877 POLY-GLN.
FT DOMAIN 1119 1128 POLY-GLY.
FT DOMAIN 1194 1198 POLY-PRO.
FT MOD_RES 99 99 PHOSPHORYLATION (BY CK2)
FT MOD_RES 460 460 PHOSPHORYLATION (BY INSR)
FT MOD_RES 608 608 PHOSPHORYLATION (BY INSR)
FT MOD_RES 628 628 PHOSPHORYLATION (BY INSR)
FT MOD_RES 891 891 PHOSPHORYLATION (BY INSR)
FT MOD_RES 935 935 PHOSPHORYLATION (BY INSR)
FT MOD_RES 983 983 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1173 1173 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1220 1220 PHOSPHORYLATION (BY INSR)
FT CONFLICT 1038 1039 MISSING (IN REF. 2).
FT CONFLICT 1182 1182 H -> R (IN REF. 2).
SQ SEQUENCE 1233 AA; 130723 MW; C0E9B2D890ADD87 CRC64;

Query Match 3.6% Score 7; DB 1; Length 1233;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 262 RSKSQSS 268
|||||||

RESULT 21
IRSI_HUMAN
ID IRS1_HUMAN STANDARD; PRT; 1242 AA.
AC P35568;
DT 01-JUN-1994 (Rel. 29, Created)

```

```

DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin receptor substrate-1 (IRS-1).
GN IRS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=93292738; PubMed=8513971;
RA Araki E., Sun X.J., Haag B.L. III, Chuang L.M., Zhang Y.,
RA Yang-Feng T.L., White M.F., Kahn C.R.;
RT "Human skeletal muscle insulin receptor substrate-1. Characterization
RT of the cDNA, gene, and chromosomal localization.";
RL Diabetes 42:1041-1054(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92181456; PubMed=1311924;
RA Nishiyama M., Wands J.R.;
RT "Cloning and increased expression of an insulin receptor substrate-1-
RT like gene in human hepatocellular carcinoma.";
RL Biochem. Biophys. Res. Commun. 183:280-285(1992).
RN [3]
RP VARIANTS PRO-512 AND ARG-971.
RX MEDLINE=93390176; PubMed=8104271;
RA Almind K., Bjoerbaek C., Vestergaard H., Hansen T., Echwald S.,
RA Pedersen O.;
RT "Aminoacid polymorphisms of insulin receptor substrate-1 in
RT non-insulin-dependent diabetes mellitus.";
RL Lancet 342:828-832(1993).
RN [4]
RP VARIANT NIDDM GLY-723 DEL.
RX MEDLINE=96303710; PubMed=8723689;
RA Esposito D.L., Mammarella S., Ranieri A., della Loggia F., Capani F.,
RA Consoli A., Mariani-Costantini R., Caramia F.G., Cama A., Battista P.;
RT "Deletion of Gly723 in the insulin receptor substrate-1 of a patient
RT with noninsulin-dependent diabetes mellitus.";
RL Hum. Mutat. 7:364-366(1996).
RN [5]
RP VARIANTS NIDDM TYR-1043 AND TYR-1095.
RA Mammarella S., Creati B., Esposito D.L., Arcuri P., della Loggia F.,
RA Capani F., Mariani-Costantini R., Caramia F.G., Battista P., Cama A.;
RT "Novel allele of the insulin receptor substrate-1 bearing two
RT non-conservative amino acid substitutions in a patient with
RT noninsulin-dependent diabetes mellitus.";
RL Hum. Mutat. 11:411-411(1998).
RN [6]
RP STRUCTURE BY NMR OF 157-267.
RX MEDLINE=96185451; PubMed=8599766;
RA Zhou M.-W., Huang B., Olejniczak E.T., Meadows R.P., Shuker S.B.,
RA Miyazaki M., Trueb T., Shoelson S.E., Fesik S.W.;
RT "Structural basis for IL-4 receptor phosphopeptide recognition by the
RT IRS-1 PTB domain.";
RL Nat. Struct. Biol. 3:388-393(1996).
CC -!- FUNCTION: MAY MEDATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC BY INSULIN. WHEN PHOSPHORYLATED BY THE INSULIN RECEPTOR BINDS
CC SPECIFICALLY TO VARIOUS CELLULAR PROTEINS CONTAINING SH2 DOMAINS
CC SUCH AS PHOSPHATIDYLINOSITOL 3-KINASE P85 SUBUNIT OR GRB-2.
CC -!- DISEASE: POLYMORPHISMS IN IRS1 MAY BE INVOLVED IN THE ETIOLOGY OF
CC A SUBSET OF LATE-ONSET NON-INSULIN-DEPENDENT DIABETES MELLITUS
CC (NIDDM).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

FT DOMAIN 272 321 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 337 383 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 432 558 CATALYTIC-A (POTENTIAL).
FT DOMAIN 578 752 CATALYTIC-B (POTENTIAL).
FT VARSPLIC 451 475 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 791 AA; 88996 MW; C7DD07F5B285FF62 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 791;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RKRSRK 134
Db 784 RKRSRK 790

RESULT 18
L100_ADEGX STANDARD; PRT; 798 AA.
AC F36856;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE Late 100 kDa protein.
OS Avian adenovirus gal10 (strain SA2) (Fowl adenovirus 10).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX NCBI_TaxID=10547;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94040780; PubMed=8224879;
RA Sheppard M.;
RT Identification of a fowl adenovirus gene with sequence homology to
RL the 100K gene of human adenovirus.;
RL Gene 132:307-308(1993).
CC -!- FUNCTION: THE 100 kDa PROTEIN IS A LATE NONSTRUCTURAL PROTEIN
CC INVOLVED IN TRANSPORT OF HEXON FROM CYTOPLASM TO THE NUCLEUS
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L07890; AAA72328.1; -
CC DR PIR; JN0878;
CC DR InterPro; IPR003381; Adeno_100.
CC DR Pfam; PF02438; adeno_100; 1.
CC KW Late protein; Transport.
CC SQ SEQUENCE 798 AA; 89028 MW; 04CFA042AC6A0DF4 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 798;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ENPNRPD 117
Db 592 ENPNRPD 598

RESULT 19
SYL_BUCAI STANDARD; PRT; 859 AA.
ID SYL_BUCAI
AC P57519;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LRSU; OR BU444.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).

```

---

```

OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Tokyo 1998.
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RL Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP001119; BAB13142.1; -
CC DR InterPro; IPR002302; Leu-TRNASyntla.
CC DR InterPro; IPR002300; tRNA-synt_1a.
CC DR InterPro; IPR001412; tRNA-synt_1.
CC DR Pfam; PF00133; tRNA-synt_1; 1.
CC DR PRINTS; PR00985; TRNASYNTHLEU.
CC DR TIGR; TIGR00396; leuS_bact; 1.
CC DR PROSITE; PS00178; AA-TRNA_LIGASE_1; 1.
CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC FT SITE 42 52 "HIGH" REGION.
CC FT SITE 618 622 "KMSKS" REGION.
CC FT BINDING 621 621 ATP (BY SIMILARITY).
CC SQ SEQUENCE 859 AA; 102170 MW; 43DB88CD4020975 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 859;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 IKEEQDK 177
Db 741 IKEEQDK 747

RESULT 20
IRSL_MOUSE STANDARD; PRT; 1233 AA.
ID IRSL_MOUSE
AC P35569;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Insulin receptor substrate-1.
DE IRSL OR IRS-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94220494; PubMed=8167159;
RA Araki E., Haag B.L., III, Kahn C.R.;
RT "Cloning of the mouse insulin receptor substrate-1 (IRS-1) gene and
RL complete sequence of mouse IRS-1."
RL Biochim. Biophys. Acta 1221:353-356(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93192326; PubMed=8448209;
RA Keller S.R., Abersold R., Garner C.W., Lienhard G.E.;
RT "The insulin-elicited 160 kDa phosphotyrosine protein in mouse
RT adipocytes is an insulin receptor substrate 1: identification by

```

DR Pfam; PF00130; DAG\_PE-bind; 2.  
DR Pfam; PF00609; DAGKa; 1.  
DR Pfam; PF00781; DAGKc; 1.  
DR PRINTS; PR00008; DAGPEDOMAIN.  
DR ProDom; PD000012; EF-hand; 1.  
DR ProDom; PD002939; DAGKa; 1.  
DR ProDom; PD005043; DAG\_kin\_cat; 1.  
DR SMART; SM00109; C1; 2.  
DR SMART; SM00045; DAGKa; 1.  
DR SMART; SM00046; DAGKc; 1.  
DR SMART; SM00054; EFh; 2.  
DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 2.  
DR PROSITE; PS0018; EF HAND; 1.  
DR Transferase; Kinase; Calcium-binding; Phorbol-ester binding;  
KW Repeat; Multigene family.  
FT CA\_BIND 185 196 EF-HAND 1 (POTENTIAL).  
FT CA\_BIND 230 241 EF-HAND 2 (POTENTIAL).  
FT DOMAIN 269 318 PHORBOL-ESTER AND DAG BINDING 1.  
FT DOMAIN 334 380 PHORBOL-ESTER AND DAG BINDING 2.  
FT DOMAIN 429 555 CATALYTIC-A (POTENTIAL).  
FT DOMAIN 575 749 CATALYTIC-B (POTENTIAL).  
SQ SEQUENCE 788 AA; 88521 MW; E77BCA40B22CID49 CRC64;  
Query Match 3.6%; Score 7; DB 1; Length 788;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 128 RRSRSK 134  
Db 781 RRSRSK 787  
RESULT 17  
KDGG\_HUMAN STANDARD; PRT; 791 AA.  
AC P49619;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Diacylglycerol kinase, gamma (EC 2.7.1.107) (Diglyceride kinase) (DGK-  
DE gamma) (DAG kinase gamma).  
CN DGK OR DAGK3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99168758; PubMed=10071200;  
RA Stoeckl H., Klein J., Gehrig A., Koehler M.R., Jurkles B., Kellner U.,  
RA Leo-Rottler B., Schmid M., Weber B.H.F.;  
RT "Mapping and genomic characterization of the gene encoding  
RT diacylglycerol kinase gamma (DAGK3): assessment of its role in  
RT dominant optic atrophy (OPAL).";  
RL Hum. Genet. 104:99-105 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99168758; PubMed=10071200;  
RA Stoeckl H., Klein J., Gehrig A., Koehler M.R., Jurkles B., Kellner U.,  
RA Leo-Rottler B., Schmid M., Weber B.H.F.;  
RT "Mapping and genomic characterization of the gene encoding  
RT diacylglycerol kinase gamma (DAGK3): assessment of its role in  
RT dominant optic atrophy (OPAL).";  
RL Hum. Genet. 104:99-105 (1999).  
CC -!- FUNCTION: REVERSES THE NORMAL FLOW OF GLYCEROLIPID  
CC BIOSYNTHESIS BY PHOSPHORYLATING DIACYLGLYCEROL BACK TO  
CC PHOSPHATIDIC ACID.  
CC -!- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-  
CC diacylglycerol 3-phosphate.  
CC -!- ENZYME REGULATION: REQUIRES PHOSPHATIDYL SERINE FOR MAXIMAL  
CC ACTIVITY.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. CAN BE LOOSELY BOUND TO THE

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM  
CC MAY BE INACTIVE.  
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN RETINA AND IN A  
CC MUCH LESSER EXTENT IN THE BRAIN. OTHER TISSUES CONTAIN EXTREMELY  
CC LOW LEVELS OF DGK-GAMMA.  
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE  
CC FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
CC BINDING DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D26135; BAA05132.1; -;  
DR EMBL; AF020945; AAC04686.1; JOINED.  
DR EMBL; AF020922; AAC04686.1; JOINED.  
DR EMBL; AF020923; AAC04686.1; JOINED.  
DR EMBL; AF020924; AAC04686.1; JOINED.  
DR EMBL; AF020925; AAC04686.1; JOINED.  
DR EMBL; AF020926; AAC04686.1; JOINED.  
DR EMBL; AF020927; AAC04686.1; JOINED.  
DR EMBL; AF020928; AAC04686.1; JOINED.  
DR EMBL; AF020929; AAC04686.1; JOINED.  
DR EMBL; AF020930; AAC04686.1; JOINED.  
DR EMBL; AF020931; AAC04686.1; JOINED.  
DR EMBL; AF020932; AAC04686.1; JOINED.  
DR EMBL; AF020933; AAC04686.1; JOINED.  
DR EMBL; AF020934; AAC04686.1; JOINED.  
DR EMBL; AF020935; AAC04686.1; JOINED.  
DR EMBL; AF020936; AAC04686.1; JOINED.  
DR EMBL; AF020937; AAC04686.1; JOINED.  
DR EMBL; AF020938; AAC04686.1; JOINED.  
DR EMBL; AF020939; AAC04686.1; JOINED.  
DR EMBL; AF020940; AAC04686.1; JOINED.  
DR EMBL; AF020941; AAC04686.1; JOINED.  
DR EMBL; AF020942; AAC04686.1; JOINED.  
DR EMBL; AF020943; AAC04686.1; JOINED.  
DR EMBL; AF020944; AAC04686.1; JOINED.  
DR Genew; HGNC:2853; DGKc.  
DR MIM; 601854; -;  
DR InterPro; IPR000756; DAGKa.  
DR InterPro; IPR001206; DAGKc.  
DR InterPro; IPR002219; DAG\_PE-bind.  
DR InterPro; IPR003622; DAG\_kin\_cat.  
DR InterPro; IPR002048; EF-hand.  
DR Pfam; PF00036; efhand; 2.  
DR Pfam; PF00130; DAG\_PE-bind; 2.  
DR Pfam; PF00609; DAGKa; 1.  
DR Pfam; PF00781; DAGKc; 1.  
DR PRINTS; PR00008; DAGPEDOMAIN.  
DR ProDom; PD000012; EF-hand; 1.  
DR ProDom; PD002939; DAGKa; 1.  
DR ProDom; PD005043; DAG\_kin\_cat; 1.  
DR SMART; SM00109; C1; 2.  
DR SMART; SM00045; DAGKa; 1.  
DR SMART; SM00046; DAGKc; 1.  
DR SMART; SM00054; EFh; 2.  
DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 2.  
DR PROSITE; PS0018; EF HAND; 2.  
KW Transferase; Kinase; Calcium-binding; Phorbol-ester binding;  
KW Repeat; Multigene family; Alternative splicing.  
FT DOMAIN 151 156 POLY-SER.  
FT CA\_BIND 188 199 EF-HAND 1 (POTENTIAL).  
FT CA\_BIND 233 244 EF-HAND 2 (POTENTIAL).

DR PROSITE; PS00372; PTS\_EI1A\_2; 1.  
KW Phosphotransferase system; Sugar transport; Transferase;  
KW Transmembrane; Phosphorylation; Complete proteome.  
FT DOMAIN 1 333 EIIC DOMAIN.  
FT DOMAIN 334 485 EIIB DOMAIN.  
FT DOMAIN 485 631 EIIB DOMAIN.  
FT TRANSMEM 20 40 POTENTIAL.  
FT TRANSMEM 52 72 POTENTIAL.  
FT TRANSMEM 86 106 POTENTIAL.  
FT TRANSMEM 134 154 POTENTIAL.  
FT TRANSMEM 161 181 POTENTIAL.  
FT TRANSMEM 213 233 POTENTIAL.  
FT TRANSMEM 269 289 POTENTIAL.  
FT TRANSMEM 313 333 POTENTIAL.  
FT MOD\_RES 256 256 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 380 380 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 548 548 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 632 AA; 69974 MW; DC413636A0BFF52 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 632;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 ILLKYDF 62  
|||||||  
Db 332 ILLKYDF 338

RESULT 15  
KGG\_MOUSE  
ID KGG\_MOUSE STANDARD; PRT; 788 AA.  
AC Q91WGT;  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE Diacylglycerol kinase, gamma (EC 2.7.1.107) (Diglyceride kinase) (DGK-gamma) (DAG kinase gamma) (88 kDa diacylglycerol kinase).  
GN DGK OR DAGK3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kidney;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol -> ADP + 1,2-diacylglycerol 3-phosphate.  
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG BINDING DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; BC015278; AAH15278.1; -  
CC MGD; MGI:105060; Dagk3.  
DR InterPro; IPR000756; DAGKa.  
DR InterPro; IPR001206; DAGKc.  
DR InterPro; IPR002219; DAG\_PE-bind.  
DR InterPro; IPR003622; DAG\_kin\_cat.  
DR Pfam; PF000036; efhand; 2.  
DR Pfam; PF00130; DAG\_PE-bind; 2.  
DR Pfam; PF00609; DAGKa; 1.

DR Pfam; PF00781; DAGKc; 1.  
DR ProDom; PD000012; EF-hand; 1.  
DR ProDom; PD002939; DAGKa; 1.  
DR ProDom; PD005043; DAG\_kin\_cat; 1.  
DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 2.  
DR PROSITE; PS00018; EF\_HAND; 2.  
KW Transferase; Kinase; Calcium-binding; Phorbol-ester binding;  
KW Repeat; Multigene family.  
FT CA\_BIND 185 196 EF-HAND 1 (POTENTIAL).  
FT CA\_BIND 230 241 EF-HAND 2 (POTENTIAL).  
FT DOMAIN 269 318 PHORBOL-ESTER AND DAG BINDING 1.  
FT DOMAIN 334 380 PHORBOL-ESTER AND DAG BINDING 2.  
FT DOMAIN 429 555 CATALYTIC-A (POTENTIAL).  
FT DOMAIN 575 749 CATALYTIC-B (POTENTIAL).  
SQ SEQUENCE 788 AA; 88523 MW; 969298DFD144F5DF CRC64;

Query Match 3.6%; Score 7; DB 1; Length 788;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKSRK 134  
|||||||  
Db 781 RRKSRK 787

RESULT 16  
KGG\_RAT  
ID KGG\_RAT STANDARD; PRT; 788 AA.  
AC P49620;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Diacylglycerol kinase, gamma (EC 2.7.1.107) (Diglyceride kinase) (DGK-gamma) (DAG kinase gamma) (88 kDa diacylglycerol kinase).  
GN DGK OR DAGK3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Wistar;  
RX MEDLINE-95108095; PubMed-7809169;  
RA Goto K., Funayama M., Kondo H.;  
RA "Cloning and expression of a cytoskeleton-associated diacylglycerol kinase that is dominantly expressed in cerebellum.";  
RL Proc Natl Acad Sci U S A. 91:13042-13046(1994).  
CC -1- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol -> ADP + 1,2-diacylglycerol 3-phosphate.  
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN BRAIN AND LOCALIZED PREDOMINANTLY IN CEREBELLAR PURKINJE CELLS.  
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG BINDING DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D38448; BAA07480.1; -  
DR InterPro; IPR000756; DAGKa.  
DR InterPro; IPR001206; DAGKc.  
DR InterPro; IPR002219; DAG\_PE-bind.  
DR InterPro; IPR003622; DAG\_kin\_cat.  
DR InterPro; IPR002048; EF-hand.  
DR Pfam; PF000036; efhand; 2.

```

Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
DB 342 RKSRKS 348

RESULT 12
NCAP_CVPRM
ID NCAP_CVPRM STANDARD; PRT; 382 AA.
AC P24411;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Nucleocapsid protein (N structural protein).
GN N.
OS Porcine respiratory coronavirus (strain RM4) (PRCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=111148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91073120; PubMed=2174956;
RA Rasschaert D., Duarte M., Laude H.;
RT "Porcine respiratory coronavirus differs from transmissible
RT gastroenteritis virus by a few genomic deletions.";
RL J. Gen. Virol. 71:2599-2607(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 224675; CAA80841.1; -
DR PIR; E36607; E36607.
DR InterPro; IPR001218; Corona_nucleocap.
DR Pfam; PF00937; Corona_nucleoca; 1.
KW Nucleocapsid.
SQ SEQUENCE 382 AA; 43551 MW; 76FE7091E5D93BC9 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 382;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
DB 342 RKSRKS 348

RESULT 13
YX8_YEAST
ID YX8_YEAST STANDARD; PRT; 577 AA.
AC P53075;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 66.9 kDa protein in SAP4-OST5 intergenic region.
GN YGL228W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Fartmann B., Kramer W.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO YEAST YPR039C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 272750; CAA96945.1; -
DR SGD; S0003197; SHE10.
KW Hypothetical protein.
SQ SEQUENCE 577 AA; 68862 MW; EF4CA95FA3157660 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 577;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 KYDFIVG 65
DB 175 KYDFIVG 181

RESULT 14
PTMA_BUCAI
ID PTMA_BUCAI STANDARD; PRT; 632 AA.
AC P57635;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE PTS system, mannitol-specific IIAC component (EIIABC-Mtl) (Mannitol-
DE permease IIAC component) (Phosphotransferase enzyme II, ABC
DE component) (EC 2.7.1.69) (EII-Mtl).
GN MTIA OR BU572.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=2045173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- PTM: AN INTRAMOLECULAR PHOSPHOTRANSFER TAKES PLACES BETWEEN
CC HIS-548 AND CYS-380 (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 PTS EIIA DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTS EIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTS EIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001119; BAB13262.1; -
DR HSSP; P00550; IA3A.
DR InterPro; IPR002178; PTS_EIIA_2.
DR Pfam; PF00359; PTS_EIIA_2; 1.
DR ProDom; PD001689; PTS_EIIA_2; 1.

```



DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-APR-1990 (Rel. 14, Last annotation update)  
 DE Nucleocapsid protein (N structural protein).  
 GN N.  
 OS Porcine transmissible gastroenteritis coronavirus (strain FS772/70).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=111150;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88216185; PubMed=2835592;  
 RA Britton P., Carmenes R.S., Page K.W., Garwes D.J., Parra F.:  
 RT "Sequence of the nucleoprotein gene from a virulent British field  
 RT isolate of transmissible gastroenteritis virus and its expression in  
 RT Saccharomyces cerevisiae.";  
 RL Mol. Microbiol. 2:89-99(1988).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; Y00542; CAA68607.1; -.  
 DR PIR; S03762; S03762.  
 DR InterPro; IPR001218; Corona\_nucleocap.  
 DR Pfam; PF00937; Corona\_nucleoca; 1.  
 KW Nucleocapsid.  
 SQ SEQUENCE 382 AA; 43483 MW; 25152E1D8F327F9E CRC64;  
 Query Match 3.6%; Score 7; DB 1; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 129 RKSRSSKS 135  
 Db 342 RKSRSSKS 348  
 RESULT 10  
 NCAP\_CVPPU STANDARD; PRT; 382 AA.  
 AC P04134;  
 DT 01-NOV-1986 (Rel. 03, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-APR-1990 (Rel. 14, Last annotation update)  
 DE Nucleocapsid protein (N structural protein).  
 GN N.  
 OS Porcine transmissible gastroenteritis coronavirus (strain Purdue).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=111151;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88078100; PubMed=2825819;  
 RA Rasschaert D., Gelfi J., Laude H.;  
 RT "Enteric coronavirus TGEV: partial sequence of the genomic RNA, its  
 RT organization and expression.";  
 RL Biochimie 69:591-600(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86181608; PubMed=3008432;  
 RA Kapke P.A., Brian D.A.;  
 RT "Sequence analysis of the porcine transmissible gastroenteritis  
 RT coronavirus nucleocapsid protein gene.";  
 RL Virology 151:41-49(1986).  
 RN [3]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RX MEDLINE=88306229; PubMed=2841792;  
 RA Kapke P.A., Tung F.Y.T., Hogue B.G., Brian D.A., Woods R.D.,  
 RA Wesley R.;

RT "The amino-terminal signal peptide on the porcine transmissible  
 RT gastroenteritis coronavirus matrix protein is not an absolute  
 RT requirement for membrane translocation and glycosylation.";  
 RL Virology 165:367-376(1988).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X06371; CAA29674.1; -.  
 DR EMBL; M14878; AAA47915.1; -.  
 DR EMBL; M21627; AAA47913.1; -.  
 DR PIR; A04025; VHIHPC.  
 DR PIR; S03936; S03936.  
 DR InterPro; IPR001218; Corona\_nucleocap.  
 DR Pfam; PF00937; Corona\_nucleoca; 1.  
 KW Nucleocapsid.  
 FT CONFLICT 237 237 W -> S (IN REF. 2).  
 FT CONFLICT 376 376 N -> I (IN REF. 2).  
 SQ SEQUENCE 382 AA; 43522 MW; E299502A0FB36EAA CRC64;  
 Query Match 3.6%; Score 7; DB 1; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 129 RKSRSSKS 135  
 Db 342 RKSRSSKS 348  
 RESULT 11  
 NCAP\_CVPR8 STANDARD; PRT; 382 AA.  
 AC P33463;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Nucleocapsid protein (N structural protein).  
 GN N.  
 OS Porcine respiratory coronavirus (strain 86/137004 / British isolate)  
 OS (PRCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=33736;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92116634; PubMed=1662846;  
 RA Britton P., Mawditt K.L., Page K.W.;  
 RT "The cloning and sequencing of the virion protein genes from a  
 RT British isolate of porcine respiratory coronavirus: comparison with  
 RT transmissible gastroenteritis virus genes.";  
 RL Virus Res. 21:181-198(1991).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X60056; CAA42657.1; -.  
 DR PIR; S24282; S24282.  
 DR InterPro; IPR001218; Corona\_nucleocap.  
 DR Pfam; PF00937; Corona\_nucleoca; 1.  
 KW Nucleocapsid.  
 SQ SEQUENCE 382 AA; 43553 MW; 50350CED86FFCD4 CRC64;  
 Query Match 3.6%; Score 7; DB 1; Length 382;

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Leivasalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConachie L.J., McTay K., McMurray A.A.,  
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM SRP55-1).  
RC TISSUE-Placenta;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 21-27 AND 47-55.  
RX MEDLINE=92249775; PubMed=1577277;  
RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;  
RT "SR proteins: a conserved family of pre-mRNA splicing factors.";  
RL Genes Dev. 6:837-847(1992).  
CC -1- FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE  
CC THE SELECTION OF ALTERNATIVE SPLICE SITES.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: SRP55-1 (SHOWN HERE), SRP55-  
CC 2 AND SRP55-3; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS  
CC DOMAIN (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).  
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U30883; AAA93073.1; -;  
DR EMBL; U30828; AAA93071.1; -;  
DR EMBL; U30829; AAA93072.1; -;  
DR EMBL; AL031681; CAB43960.1; -;  
DR EMBL; BC006832; AAH06832.1; -;  
DR Genew; HgNC:10788; SFRS6.  
DR MIM; 601944; -;  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 2.  
DR SMART; SM00360; rrm; 2.  
DR PROSITE; PS50102; RRM; 2.  
DR PROSITE; PS00030; RRM\_LNP\_1; FALSE\_NEG.  
KW Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;  
KW Repeat; Phosphorylation.  
FT DOMAIN 1 72 RNA-BINDING (RRM) 1.  
FT DOMAIN 87 90 GLY-RICH (HINGE REGION).  
FT DOMAIN 110 183 RNA-BINDING (RRM) 2.  
FT DOMAIN 184 343 ARG/SER-RICH (RS DOMAIN).  
FT VARSPLIC 86 135 SGGGGYSRRKTSRGGPPVTEYRLIVLNLSSRSWODL  
FT KDFMRQAGE -> MTNGAEAVSTEARMTAFDPWPLFLTLC

FT DPCPMTLWLTLPAMTTAAFC (IN ISOFORM SRP55-  
FT 2).  
FT MISSING (IN ISOFORM SRP55-2).  
FT RVSPPPKRATSRSRSRSRSSSRD -> LKLGCA  
FT RFMSQQTESLISLASC (IN ISOFORM SRP55-3).  
FT R -> H (IN REF. 1).  
SQ SEQUENCE 344 AA: 39586 MW: 72305506CE948B94 CRC64;  
Query Match 3.6%; Score 7; DB 1; Length 344;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 129 RKSRKS 135  
DB 241 RKSRKS 247  
|||||||  
RESULT 8  
DHPL\_BACSH STANDARD; PRT; 381 AA.  
AC P23307;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phenylalanine dehydrogenase (EC 1.4.1.20) (PheDH).  
GN PDH.  
OS Bacillus sphaericus.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1421;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88255879; PubMed=2838396;  
RA Okazaki N., Hibino Y., Asano Y., Ohmori M., Numao N., Kondo K.;  
RT "Cloning and nucleotide sequencing of phenylalanine dehydrogenase  
RL gene of Bacillus sphaericus.";  
RL Gene 63:337-341(1988).  
CC -1- CATALYTIC ACTIVITY: L-phenylalanine + H(2)O + NAD(+) =  
CC phenylpyruvate + NH(3) + NADH.  
CC -1- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M26661; AAA22646.1; -;  
DR InterPro; IPR001625; GLFV.Dh.  
DR Pfam; PF00208; GLFV\_dehydrog; 1.  
DR Pfam; PF02812; GLFV\_dehydrog\_N; 1.  
DR PRINTS; PR00082; GLFVDEHGRNASE.  
DR PROSITE; PS00074; GLFV\_DEHYDROGENASE; 1.  
DR KX Oxidoreductase; NAD.  
FT ACT\_SITE 91 BY SIMILARITY.  
FT NP\_BIND 191 197 NAD (POTENTIAL).  
SQ SEQUENCE 381 AA: 41578 MW: D32450CE397ACA67 CRC64; ;  
Query Match 3.6%; Score 7; DB 1; Length 381;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 26 IVKSDDI 32  
DB 236 IVKSDDI 242  
|||||||  
RESULT 9  
NCAP\_CVPFS STANDARD; PRT; 382 AA.  
ID NCAP\_CVPFS  
AC P05991;  
DT 01-NOV-1988 (Rel. 09, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sec-independent protein translocase protein tatB.  
 GN TATB OR HI0187.1.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitchugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 CC -|- FUNCTION: REQUIRED FOR CORRECT LOCALIZATION OF PRECURSOR PROTEINS  
 CC BEARING SIGNAL PEPTIDES WITH THE TWIN ARGININE CONSERVED MOTIF  
 CC S/T-R-X-F-L-K. THIS SEC-INDEPENDENT PATHWAY IS TERMED TAT FOR  
 CC TWIN-ARGININE TRANSLLOCATION SYSTEM. THIS SYSTEM MAINLY TRANSPORTS  
 CC PROTEINS WITH BOUND COFACTORS THAT REQUIRE FOLDING PRIOR TO EXPORT  
 CC (BY SIMILARITY).  
 CC -|- SUBCELLULAR LOCATION: INNER-MEMBRANE BOUND (PROBABLE).  
 CC -|- SIMILARITY: BELONGS TO THE TATB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U32704; -; NOT\_ANNOTATED\_CDS.  
 DR TIGR: HI0187.1; -;  
 DR InterPro: IPR003998; TatB.  
 DR PRINTS; PR01506; TATBPROTEIN.  
 DR TIGRFAMS; TIGR01014; 2a6401s01; 1.  
 DR Transport; Protein transport; Translocation; Transmembrane;  
 KW Inner membrane; Complete proteome.  
 FT TRANSMEM 1 21 POTENTIAL.  
 SQ SEQUENCE 186 AA; 20430 MW; 7FD2166488CA057F CRC64;  
 Query Match 3.6%; Score 7; DB 1; Length 186;  
 Best Local Similarity 100.0%; Pred. No. 6.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 133 SKSQSSK 139  
 Db 177 SKSQSSK 183  
 |||||  
 RESULT 6  
 RACH\_DICDI STANDARD; PRT; 200 AA.  
 ID RACH\_DICDI  
 AC Q9GPR7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE RAS-related protein rach.  
 GN RACH.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.  
 OX NCBI\_TaxID=44689;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RX MEDLINE=21127961; PubMed=11222756;  
 RA Rivero F., Dislich H., Glockner G., Noegel A.A.;  
 RT "The Dictyostelium discoideum family of Rho-related proteins.";  
 RL Nucleic Acids Res. 29:1068-1079(2001).  
 CC -|- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF310894; AAG45133.1; -;  
 DR HSP; P21181; IAMA4.  
 DR DictyDB; DD?????; rach.  
 DR InterPro: IPR003578; GTPase\_Rho.  
 DR InterPro: IPR001230; Prenyl\_site.  
 DR InterPro: IPR001806; Ras\_trnsmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS; PR00449; RASTRNSPRMG.  
 DR SMART; SM00174; RHO; 1.  
 DR TIGRFAMS; TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Prenylation; Lipoprotein.  
 FT NP\_BIND 11 18 GTP (BY SIMILARITY).  
 FT NP\_BIND 58 62 GTP (BY SIMILARITY).  
 FT NP\_BIND 117 120 GTP (BY SIMILARITY).  
 FT DOMAIN 33 41 EFFECTOR REGION (POTENTIAL).  
 FT LIPID 197 197 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 200 AA; 22189 MW; 0852F4358EC6A648 CRC64;  
 Query Match 3.6%; Score 7; DB 1; Length 200;  
 Best Local Similarity 100.0%; Pred. No. 6.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 KEVTPEM 9  
 Db 130 KEVTPEM 136  
 |||||  
 RESULT 7  
 SFR6\_HUMAN STANDARD; PRT; 344 AA.  
 ID SFR6\_HUMAN  
 AC Q13247; Q13244; Q13245; Q9UJB8; Q96J06;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Splicing factor, arginine/serine-rich 6 (pre-mRNA splicing factor  
 DE SRP55).  
 GN SFRS6 OR SRP55.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE-Colon;  
 RX MEDLINE=96016206; PubMed=7556075;  
 RA Screaton G.R., Caceres J.F., Mayeda A., Bell M.V., Plebanski M.,  
 RA Jackson D.G., Bell J.I., Krainer A.R.;  
 RT "Identification and characterization of three members of the human SR  
 RT family of pre-mRNA splicing factors.";  
 RL EMBO J. 14:4336-4349(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beare D.M.,

```
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L32P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL021999; -; NOT ANNOTATED CDS.
DR EMBL: AE006985; AAK45255.1; ALT_INIT.
DR TIGR: MT1007; -.
DR TubercuList; Rv0979.lc; -.
DR InterPro; IPR002677; Ribosomal_L32p.
DR Pfam; PF01783; Ribosomal_L32p; 1.
DR Ribosomal protein; Complete proteome.
KW RIBOSOMAL
SQ SEQUENCE 57 AA; 6507 MW; 2EF45101DDBF5921 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 KRRKRS 133
Db 5 KRRKRS 11

RESULT 3
SFR6_RABIT STANDARD; PRT; 81 AA.
ID SFR6_RABIT
AC O18776;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Splicing factor, arginine/serine-rich 6 (Pre-mRNA splicing factor
DE SRP55) (Fragment).
DE SRP56 OR SRP55.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCB1_TaxID-9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-New Zealand white;
RA Brunet A., Henrion G., Duranthon V., Renard J.P.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE
CC THE SELECTION OF ALTERNATIVE SPLICE SITES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL: AF011564; AAB66467.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR PROSITE; PS50102; RRM; PARTIAL.
DR PROSITE; PS00030; RRM_RNP_1; PARTIAL.
KW Nuclear protein; RNA-binding; mRNA splicing; Repeat; Phosphorylation.
FT NON_TER 1
FT DOMAIN <1 >81 ARG/SER-RICH (RS DOMAIN).
FT NON_TER 81
SQ SEQUENCE 81 AA; 9520 MW; 20D9F8A9D08C7CA0 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 81;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
Db 39 RKSRKS 45

RESULT 4
RR12_PORPU STANDARD; PRT; 124 AA.
ID RR12_PORPU
AC P51289;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Chloroplast 30S ribosomal protein S12.
GN RPS12.
OS Porphyra purpurea.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCB1_TaxID-2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Avonport;
RA Reith M.E., Munholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -1- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U38804; AAC08175.1; -.
DR InterPro; IPR000230; Ribosomal_S12.
DR Pfam; PF00164; Ribosomal_S12; 1.
DR PRINTS; PR01034; RIBOSOMALS12.
DR PRODOM; PD000576; Ribosomal_S12; 1.
DR TIGRFAMS; TIGR00981; rpsL_bact; 1.
DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 124 AA; 13904 MW; 6632BE623C11FB1 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 RKSRSK 134
Db 110 RKSRSK 116

RESULT 5
TATB_HAETN STANDARD; PRT; 186 AA.
ID TATB_HAETN
AC P57047; P44559;
DT 01-NOV-1995 (Rel. 32, Created)
```

983 5 2.6 280 1 Y705\_CHLPP Q927k0 chlamydia p  
 984 5 2.6 280 1 YB00\_YEAST P32788 saccharomyc  
 985 2.6 281 1 CLCX\_CAEEL Q9ngj7 caenorhabdi  
 986 5 2.6 281 1 MSP1\_PLAFN P13820 plasmodium  
 987 5 2.6 281 1 NPS2\_MOUSE O55126 mus musculus  
 988 5 2.6 281 1 TNF6\_HUMAN P48023 homo sapien  
 989 5 2.6 282 1 ARG6\_THEMA Q92a24 thermotoga  
 990 5 2.6 282 1 LPXC\_CHLPP Q927q2 chlamydia p  
 991 5 2.6 282 1 TR2A\_HUMAN Q13595 homo sapien  
 992 5 2.6 283 1 COBD\_PYRHO O58114 pyrococcus  
 993 5 2.6 283 1 FNKC\_ECOLI P14190 escherichia  
 994 5 2.6 283 1 IPF1\_HUMAN P52945 h insulin p  
 995 5 2.6 283 1 IPF1\_MESAU P70118 mesocricetu  
 996 5 2.6 283 1 IPF1\_RAT P52947 rattus norv  
 997 5 2.6 283 1 MTD\_METKA P94951 methanopyru  
 998 5 2.6 283 1 SFRA\_MOUSE Q60701 mus musculus  
 999 5 2.6 283 1 YBAS\_BACSU P55190 bacillus su  
 1000 5 2.6 284 1 IPF1\_MOUSE P52946 mus musculus

## ALIGNMENTS

RESULT 1  
 KG62\_HUMAN  
 ID KG62\_HUMAN STANDARD; PRT; 733 AA.  
 AC Q15349; Q15419; Q90JN5;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ribosomal protein S6 kinase alpha 2 (EC 2.7.1.-) (S6K-alpha 2) (90  
 DE kda ribosomal protein S6 kinase 2 (EC 2.7.1.-) (Ribosomal S6 kinase 3)  
 DE (RSK-3) (pp90RSK3).  
 GN RPS6KA2 OR RSK3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=95349602; PubMed=7623830;  
 RA Zhao Y., Bjoerbaek C., Weremowicz S., Morton C.C., Moller D.E.;  
 RT "RSK3 encodes a novel pp90rsk isoform with a unique N-terminal  
 RT sequence: growth factor-stimulated kinase function and nuclear  
 RT translocation.";  
 RL Mol. Cell. Biol. 15:4353-4363(1995).  
 [2]  
 RN SEQUENCE OF 1-540 FROM N.A.  
 RP MEDLINE=94189676; PubMed=8141249;  
 RA Moller D.E., Xia C.-H., Tang W., Zhu A.X., Jakubowski M.;  
 RT "Human rsk isoforms: cloning and characterization of tissue-specific  
 RT expression.";  
 RL Am. J. Physiol. 266:C351-C359(1994).  
 [3]  
 RN SEQUENCE OF 101-733 FROM N.A.  
 RP Palmer S.;  
 RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RL FUNCTION: PHOSPHORYLATES A WIDE RANGE OF SUBSTRATES INCLUDING  
 CC RIBOSOMAL PROTEIN S6. IMPLICATED IN THE ACTIVATION OF THE MITOGEN-  
 CC ACTIVATED KINASE CASCADE.  
 CC  
 CC SUBCELLULAR LOCATION: Nuclear.  
 CC TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. HIGHEST EXPRESSION  
 CC IN LUNG AND SKELETAL MUSCLE.  
 CC PTM: AUTOPHOSPHORYLATED.  
 CC SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC S6 KINASE SUBFAMILY. CONTAINS TWO KINASE DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

DR EMBL; X85106; CAA59427.1; -  
 DR EMBL; L07598; AAC82496.1; ALT\_INIT.  
 DR EMBL; AL022069; CAB58228.1; -  
 DR HSSP; Q63450; 1A06.  
 DR Genew; HGNC:10431; RPS6KA2.  
 DR MIM; 601685; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR000961; Pkinase-C.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; pkinase; 2.  
 DR Pfam; PF00433; pkinase\_C; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 2.  
 DR SMART; SM00133; S\_TK\_X; 1.  
 DR SMART; SM00220; S\_TK; 2.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 2.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 2.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 2.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Repeat; Multigene family; Phosphorylation; Nuclear protein.  
 FT DOMAIN 59 318  
 FT PROTEIN KINASE 1.  
 FT DOMAIN 415 672  
 FT PROTEIN KINASE 2.  
 FT NP\_BIND 65 73  
 FT BINDING 91 91  
 FT ACT\_SITE 184 184  
 FT NP\_BIND 421 429  
 FT BINDING 444 444  
 FT ACT\_SITE 532 532  
 FT CONFLICT 256 256  
 FT CONFLICT 269 269  
 FT CONFLICT 339 339  
 FT CONFLICT 447 447  
 FT CONFLICT D -> G (IN REF. 2).  
 SQ SEQUENCE 733 AA; 83252 MW; FC82095707F95033 CRC64;

Query Match 4.2%; Score 8; DB 1; Length 733;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSKS 135  
 Db 19 RRKRSKS 26  
 |||||

RESULT 2  
 RL32\_MYCTU STANDARD; PRT; 57 AA.  
 ID RL32\_MYCTU  
 AC P58287;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 50S ribosomal protein L32.  
 GN RPMF OR RV0979.1C OR MT1007 OR MTV044.07c.1.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 [2]

837	5	2.6	251	1	PROB_CAMJE	Q9p129	campylobact	910	5	2.6	264	1	YLPK_CAEEL	P91301	caenorhabdi
838	5	2.6	251	1	RSOA_YEAST	P32905	saccharomyc	911	5	2.6	264	1	YNV6_YEAST	P53853	saccharomyc
839	5	2.6	251	1	RSOB_YEAST	P46654	saccharomyc	912	5	2.6	265	1	EFID_XENLA	P29693	xenopus lae
840	5	2.6	251	1	YQRE_CAEEL	Q09308	caenorhabdi	913	5	2.6	265	1	IOD2_HUMAN	Q92813	homo sapien
841	5	2.6	252	1	1431_SCHMA	Q26540	schistosoma	914	5	2.6	265	1	PPNK_BACHD	Q96904	salmonella
842	5	2.6	252	1	PMW_CANAL	P31353	candida alb	915	5	2.6	265	1	RFAP_SALTY	Q06995	salmonella
843	5	2.6	252	1	UNG_HSVSA	Q01019	herpesvirus	916	5	2.6	265	1	RPIA_ARATH	Q92u38	arabidopsis
844	5	2.6	253	1	PHOT_LISIN	Q92dm1	listeria in	917	5	2.6	265	1	UCRI_SOLTU	P37841	solanan tub
845	5	2.6	253	1	PHOT_LISMO	P58724	listeria mo	918	5	2.6	265	1	Y280_MYCGE	P47522	mycoplasma
846	5	2.6	253	1	RSO_CANAL	Q42817	candida alb	919	5	2.6	265	1	YG09_STRPN	Q97PK0	streptococc
847	5	2.6	253	1	RS3_ABEWE	P79891	ambystoma m	920	5	2.6	266	1	AROE_HELPJ	Q92jx8	helicobacte
848	5	2.6	253	1	SOJ_BACHD	Q9K5n0	bacillus ha	921	5	2.6	267	1	DPM1_YEAST	P14020	saccharomyc
849	5	2.6	253	1	TRPA_LACLA	Q01997	lactococcus	922	5	2.6	267	1	PROB_STRTR	P96488	streptococc
850	5	2.6	253	1	UPPS_CHLTR	O84456	chlamydia t	923	5	2.6	268	1	IL2A_CANEA	O62802	canis famil
851	5	2.6	253	1	YSO2_ACTAM	P29086	acidianus a	924	5	2.6	268	1	NO30_MEDTR	P93329	medicago tr
852	5	2.6	254	1	IDNO_ECOLI	P39345	escherichia	925	5	2.6	268	1	UCR3_TOBAC	P51133	nicotiana t
853	5	2.6	254	1	YC43_PORPU	P51264	porphyra pu	926	5	2.6	268	1	UCR5_TOBAC	P51135	nicotiana t
854	5	2.6	254	1	YQCI_BACSU	P45944	bacillus su	927	5	2.6	269	1	UCRS_TOBAC	P70714	actinobacil
855	5	2.6	255	1	DLX1_MOUSE	Q64317	mus musculus	928	5	2.6	269	1	CYSQ_ACTAC	Q09167	rattus norv
856	5	2.6	255	1	SOLA_YEAST	P53315	saccharomyc	929	5	2.6	269	1	SFR5_RAT	P47758	mus musculus
857	5	2.6	256	1	CFL_ELAUM	O65333	elaeagnus u	930	5	2.6	269	1	SRPB_MOUSE	O57590	fugu rubrip
858	5	2.6	256	1	FBP2_DROME	P53378	drosophila	931	5	2.6	269	1	SUR4_FUGRU	O15360	homo sapien
859	5	2.6	256	1	M3R_DICDI	P11872	dictyosteli	932	5	2.6	269	1	SUR4_MOUSE	Q64310	mus musculus
860	5	2.6	256	1	OMAT_HUMAN	Q16633	homo sapien	933	5	2.6	269	1	VG51_HSVSA	Q01036	herpesvirus
861	5	2.6	256	1	VFAT_HRSVA	P03419	human respi	934	5	2.6	270	1	ALYS_BPRI1	Q38135	bacterioph
862	5	2.6	256	1	YREC_SYNP2	P19737	synchococc	935	5	2.6	270	1	CFWG_SCHPO	O97c5	schizosacch
863	5	2.6	257	1	AAPR_RHILV	O52815	rhizobium l	936	5	2.6	270	1	ES1_BRARE	Q90257	brachydanio
864	5	2.6	257	1	FAEL_ECOLI	P33784	escherichia	937	5	2.6	270	1	HRF5_MOUSE	Q9cpx5	mus musculus
865	5	2.6	257	1	LIVG_METJA	Q58863	methanococc	938	5	2.6	270	1	RSP4_DROME	P38979	drosophila
866	5	2.6	257	1	YBE3_YEAST	P38195	saccharomyc	939	5	2.6	270	1	SFR5_MOUSE	O35326	mus musculus
867	5	2.6	257	1	YL48_MYCLE	Q9cce2	mycobacteri	940	5	2.6	270	1	SUR4_DROME	O18405	drosophila
868	5	2.6	258	1	DTC_MOUSE	P01881	mus musculus	941	5	2.6	270	1	Y041_CHLTR	O84045	chlamydia t
869	5	2.6	258	1	LPXA_PSEAE	Q9X5P4	pseudomonas	942	5	2.6	270	1	YD1J_SCHPO	O10247	schizosacch
870	5	2.6	258	1	UCRI_TOBAC	P47929	nicotiana t	943	5	2.6	270	1	YI19_PSEAE	Q9hy16	pseudomonas
871	5	2.6	258	1	VSP2_AGRKH	P47797	agkistrodon	944	5	2.6	271	1	HMB_XENLA	P14837	xenopus lae
872	5	2.6	259	1	GUB_BACBR	P37073	bacillus br	945	5	2.6	271	1	KKAI_ECOLI	P00551	escherichia
873	5	2.6	259	1	PEB1_CAMJE	P45678	campylobact	946	5	2.6	271	1	KKAI_SALTY	Q03447	salmonella
874	5	2.6	259	1	PSTB_PASMU	Q9cnj7	pasteurella	947	5	2.6	271	1	MIND_THEMA	Q9x2i3	thermotoga
875	5	2.6	259	1	PYRK_BACHD	Q9K9u0	bacillus ha	948	5	2.6	271	1	NIE1_HUMAN	Q12981	homo sapien
876	5	2.6	259	1	YRPD_ECOLI	P45753	escherichia	949	5	2.6	271	1	SPED_CLOPE	Q8xmz0	clostridium
877	5	2.6	260	1	COX3_RAT	P05505	rattus norv	950	5	2.6	271	1	SRPB_HUMAN	Q9Y5m8	homo sapien
878	5	2.6	260	1	COX3_XENLA	P00419	xenopus lae	951	5	2.6	271	1	YHE9_YEAST	P38722	saccharomyc
879	5	2.6	260	1	DHBB_MOUSE	O00171	mus musculus	952	5	2.6	272	1	UCR2_TOBAC	Q13243	homo sapien
880	5	2.6	260	1	STCE_EMENI	Q00674	emericeila	953	5	2.6	272	1	MP74_YEAST	P51132	nicotiana t
881	5	2.6	260	1	VSP1_AGKAC	Q9i8x2	agkistrodon	954	5	2.6	273	1	PROB_STRPY	P39015	saccharomyc
882	5	2.6	260	1	VSPB_TRIGA	O13061	trimeresuru	955	5	2.6	273	1	TCIA_CAEER	Q99j17	streptococc
883	5	2.6	261	1	COX3_HUMAN	P00414	homo sapien	956	5	2.6	273	1	UCRI_MAIZE	P35072	caenorhabdi
884	5	2.6	261	1	COX3_HYLLA	Q95707	hylobates l	957	5	2.6	273	1	COAT_RBDV	P49727	zea mays
885	5	2.6	261	1	COX3_MOUSE	P00416	mus musculus	958	5	2.6	274	1	COX3_ALLMA	P23629	raspberry b
886	5	2.6	261	1	COX3_PETMA	Q35539	petromyzon	959	5	2.6	274	1	EP1Z_STAEP	P80439	allomyces m
887	5	2.6	261	1	COX3_PONPA	P92696	pongo pygma	960	5	2.6	275	1	PH4H_RHILO	P30193	staphylococ
888	5	2.6	261	1	VE39_NPVOP	Q05121	orgyia pseu	961	5	2.6	275	1	VA16_VACCV	Q98472	rhizobium l
889	5	2.6	262	1	FBOL_BOMMO	P21828	bombyx mori	962	5	2.6	275	1	YRB2_SYNP7	P16710	vaccinia vi
890	5	2.6	262	1	FRGI_DROME	Q9vwa8	drosophila	963	5	2.6	275	1	PTP3_CHLEU	P46205	synchococc
891	5	2.6	262	1	IOD2_MOUSE	Q9z1y9	mus musculus	964	5	2.6	276	1	PYRF_CHLEU	Q39491	chlamydomon
892	5	2.6	262	1	IOD2_RAT	P70551	rattus norv	965	5	2.6	277	1	PYRF_ASPNG	P07817	aspergillus
893	5	2.6	262	1	NCAP_INSVN	P28975	impatiens n	966	5	2.6	277	1	PYRF_ASPOR	O13416	aspergillus
894	5	2.6	262	1	NCAP_INSVN	Q01808	impatiens n	967	5	2.6	278	1	CRTC_RHOSH	O01670	rhodobacter
895	5	2.6	262	1	RNC_MYCGE	P47607	mycoplasma	968	5	2.6	278	1	EXSA_PSEAE	P26993	pseudomonas
896	5	2.6	262	1	YBEM_ECO57	P58054	escherichia	969	5	2.6	278	1	OSTP_BOVIN	P31096	bos taurus
897	5	2.6	262	1	YBEM_ECOLI	P39874	escherichia	970	5	2.6	278	1	PYRF_ASPFU	O13410	aspergillus
898	5	2.6	263	1	AROE_HELPJ	P56119	helicobacte	971	5	2.6	278	1	PYRF_CLAFU	Q9hfv8	cladosporiu
899	5	2.6	263	1	Y169_METJA	Q57633	methanococc	972	5	2.6	278	1	PYRF_PENCH	P09463	penicillium
900	5	2.6	263	1	YBGF_ECOLI	P45955	escherichia	973	5	2.6	278	1	TNF6_RAT	P36940	rattus norv
901	5	2.6	263	1	YHCK_ECOLI	P45427	escherichia	974	5	2.6	279	1	DHPI_ECOLI	P11744	escherichia
902	5	2.6	264	1	CXBI_XENLA	P08983	xenopus lae	975	5	2.6	279	1	HRF5_HUMAN	Q96kn8	homo sapien
903	5	2.6	264	1	G3P_BORHE	P46796	borrelia he	976	5	2.6	279	1	TNF6_MOUSE	P41047	mus musculus
904	5	2.6	264	1	HRCA_CAMJE	Q9ppg2	campylobact	977	5	2.6	279	1	Y43D_MYCPN	P75155	mycoplasma
905	5	2.6	264	1	RS2_HELPJ	Q9z170	helicobacte	978	5	2.6	280	1	CX32_ARATH	P27450	arabidopsis
906	5	2.6	264	1	RS2_HELPJ	P56009	helicobacte	979	5	2.6	280	1	NADE_AERPE	Q9yail	aeropyrum p
907	5	2.6	264	1	RSF4_STRPU	P46771	strongyloce	980	5	2.6	280	1	TNF6_CERTO	Q9bnd1	cercobace
908	5	2.6	264	1	WBBI_ECOLI	P36667	escherichia	981	5	2.6	280	1	TNF6_MACMU	Q9my16	macaca mula
909	5	2.6	264	1	Y557_PASMU	Q9cn81	pasteurella	982	5	2.6	280	1	Y466_HAEIN	P44000	haemophilus

691	5	2.6	219	1	DKL1_FOWPV	P21974	fowlpox vir	764	5	2.6	233	1	MRKB_KLEPN	P21646	klebsiella
692	5	2.6	219	1	GL22_ARATH	Q9f227	rhinodopsis	785	5	2.6	233	1	MSP1_PLAFD	P1827	plasmodium
693	5	2.6	219	1	GPH_RHOCA	Q33512	rhodobacter	786	5	2.6	234	1	KDSB_AQUAE	O66914	aquifex aeo
694	5	2.6	219	1	GSHE_PIG	O18994	sus scrofa	767	5	2.6	234	1	YSP1_RHIME	Q92sn8	rhizobium m
695	5	2.6	219	1	PA44_ECOLI	P22996	escherichia	768	5	2.6	234	1	YSP1_AGRKH	P26324	agkistrodon a
696	5	2.6	219	1	UNG_LACIA	Q9cix2	lactococcus	769	5	2.6	234	1	YGG5_ECOLI	P52054	escherichia
697	5	2.6	220	1	AROD_METJA	Q58849	methanococc	770	5	2.6	234	1	YKFA_BACSU	O34851	bacillus su
698	5	2.6	220	1	NRG2_YEAST	P38082	saccharomyc	771	5	2.6	235	1	1433_XENLA	P29309	xenopus lae
699	5	2.6	220	1	ROX3_YEAST	P25046	saccharomyc	772	5	2.6	235	1	MSRA_HOMAB	Q9u168	homo sapien
700	5	2.6	220	1	YICG_HAEIN	P45122	haemophilus	773	5	2.6	235	1	RPSE_CLOAB	P33657	clostridium
701	5	2.6	221	1	GSHP_RAT	Q64625	rattus norv	774	5	2.6	235	1	YIF2_YEAST	P40455	saccharomyc
702	5	2.6	221	1	GSPJ_VIBCH	P45776	vibrio chol	775	5	2.6	236	1	COAT_MCMV	P11642	maize chlor
703	5	2.6	221	1	MLTE_BUCAI	P57352	buchnera ap	776	5	2.6	236	1	RS5_PYRAB	O9v1v5	pyrococcus
704	5	2.6	221	1	NUSB_SYNV3	P74395	synecocyst	777	5	2.6	236	1	RS5_PYRHO	O59439	pyrococcus
705	5	2.6	221	1	SEGA_BP74	P32286	bacterioph	778	5	2.6	236	1	UCR4_TOBAC	P51134	nicotiana t
706	5	2.6	221	1	SFR2_CHICK	P30352	gallus gall	779	5	2.6	236	1	YMI3_CAEEL	Q21018	caenorhabdi
707	5	2.6	221	1	SFR2_HUMAN	Q01130	homo sapien	780	5	2.6	237	1	YMI3_RHELO	Q98dd5	rhizobium l
708	5	2.6	221	1	SFR2_MOUSE	Q62093	mus musculus	781	5	2.6	237	1	SPIN_HUMAN	Q9y657	homo sapien
709	5	2.6	222	1	FLGH_RHOSH	P58205	rhodobacter	782	5	2.6	238	1	IBP6_MOUSE	P47880	mus musculus
710	5	2.6	222	1	HXC5_HUMAN	Q00444	homo sapien	783	5	2.6	238	1	RK5_ODOSI	P49547	odontella s
711	5	2.6	222	1	HXC5_MOUSE	P32043	mus musculus	784	5	2.6	238	1	SPR7_HUMAN	Q16629	homo sapien
712	5	2.6	222	1	YEOT_SCHPO	O13785	schizosacch	785	5	2.6	239	1	OMPR_ECOLI	P03025	escherichia
713	5	2.6	223	1	CDRI_HUMAN	P51861	homo sapien	786	5	2.6	239	1	RP35_BACTK	P26763	bacillus th
714	5	2.6	223	1	PATM_VIBHA	P52625	vibrio harv	787	5	2.6	239	1	RPSE_BACSU	P06222	bacillus su
715	5	2.6	223	1	RB21_CANFA	P95745	canis fami	788	5	2.6	239	1	RS3_SYNV3	P73314	synecocyst
716	5	2.6	223	1	UBL1_HORSE	Q9gm50	equus cabal	789	5	2.6	239	1	RS3_YEAST	P05750	saccharomyc
717	5	2.6	223	1	UBL1_HUMAN	P09936	homo sapien	790	5	2.6	239	1	T2E8_ECOLI	P50195	escherichia
718	5	2.6	223	1	UBL1_MOUSE	Q9r0p9	mus musculus	791	5	2.6	239	1	Y332_MYCGE	P47574	mycoplasma
719	5	2.6	223	1	UBL1_RAT	Q00981	rattus norv	792	5	2.6	240	1	CC14_SCHPO	P36389	schizosacch
720	5	2.6	224	1	YF94_HAEIN	P44265	haemophilus	793	5	2.6	240	1	LPXH_ECO57	Q8xcul	escherichia
721	5	2.6	224	1	Y209_LISMO	P33381	listeria mo	794	5	2.6	240	1	LPXH_ECOLI	P4341	escherichia
722	5	2.6	224	1	Y241_LISIN	Q92f66	listeria in	795	5	2.6	240	1	SPIN_MOUSE	O61142	mus musculus
723	5	2.6	224	1	YJ07_YEAST	P47056	saccharomyc	796	5	2.6	240	1	YD37_MYCTU	Q10647	mycobacteri
724	5	2.6	225	1	COX2_APILI	P20375	apis mellif	797	5	2.6	241	1	CLII_HUMAN	O00299	homo sapien
725	5	2.6	225	1	RB21_HUMAN	Q9u125	homo sapien	798	5	2.6	241	1	CLII_MOUSE	Q92zlg5	mus musculus
726	5	2.6	225	1	VL73_FOWPV	Q9j566	fowlpox vir	799	5	2.6	241	1	YFHQ_HAEIN	P44676	haemophilus
727	5	2.6	226	1	DCTR_BACSU	P96602	bacillus su	800	5	2.6	242	1	TRPD_BACCA	P30525	bacillus ca
728	5	2.6	226	1	IBP6_RAT	P35572	rattus norv	801	5	2.6	242	1	YD26_HAEIN	P44162	haemophilus
729	5	2.6	226	1	TPIS_METKA	Q8tut9	methanopyru	802	5	2.6	243	1	CAVT_BRALA	P05548	branchiost
730	5	2.6	226	1	TRPF_METTM	P26941	methanobact	803	5	2.6	243	1	RS3_HUMAN	P23396	homo sapien
731	5	2.6	226	1	YQ00_LISIN	Q928d5	listeria in	804	5	2.6	243	1	RS3_MANSE	P48153	manduca sex
732	5	2.6	226	1	YRAL_YEAST	Q12159	saccharomyc	805	5	2.6	243	1	RS3_MOUSE	P17073	mus musculus
733	5	2.6	226	1	YS21_CABEL	Q09652	caenorhabdi	806	5	2.6	244	1	BOS1_YEAST	P25385	saccharomyc
734	5	2.6	227	1	ID1L_SCHPO	Q10132	schizosacch	807	5	2.6	244	1	RS4E_SULAC	O05634	sulfolobus
735	5	2.6	227	1	PHLB_SERLI	P18954	serratia li	808	5	2.6	245	1	KSU5_ECOLI	O05634	sulfolobus
736	5	2.6	227	1	RMJ_RICPR	Q92e00	rickettsia	809	5	2.6	245	1	RS3_ICTPU	P02416	escherichia
737	5	2.6	228	1	CAPB_STRAU	P39851	staphylococ	810	5	2.6	246	1	BRM1_HUMAN	Q90y82	ictalurus p
738	5	2.6	228	1	IFE2_CAEEL	Q21693	caenorhabdi	811	5	2.6	246	1	BRM1_MOUSE	Q9hcu9	homo sapien
739	5	2.6	228	1	LCK_AEDAE	Q03962	aedes aegypt	812	5	2.6	246	1	Y586_BUCAE	Q99n20	mus musculus
740	5	2.6	228	1	PCAI_ACICA	O43973	acinetobact	813	5	2.6	246	1	NO25_MEDSA	P10322	medicago sa
741	5	2.6	228	1	TF2D_PLAFA	P32086	plasmodium	814	5	2.6	246	1	RIMM_ANASP	O8y1b1	anabaena sp
742	5	2.6	228	1	VSPA_LACMU	P33589	lachesis mu	815	5	2.6	246	1	RS3A_XENLA	P02350	xenopus lae
743	5	2.6	228	1	X855_METJA	Q58265	methanococc	816	5	2.6	246	1	RS3B_XENLA	P47935	xenopus lae
744	5	2.6	229	1	R33B_HUMAN	Q9H082	homo sapien	817	5	2.6	246	1	Y586_BUCAE	Q06559	drosophila
745	5	2.6	229	1	R33B_MOUSE	Q35963	mus musculus	818	5	2.6	246	1	YSW4_CAEEL	P57646	buchnera ap
746	5	2.6	229	1	RIBB_METTH	O27543	methanobact	819	5	2.6	247	1	PCNL_AERPE	Q10019	caenorhabdi
747	5	2.6	229	1	SOML_TETPU	Q919h4	tetraodon m	820	5	2.6	247	1	PCNA_HALMA	Q9yft8	aeropyrum p
748	5	2.6	229	1	VG14_BPPH8	P14814	bacterioph	821	5	2.6	247	1	PCNA_HALNI	Q9h445	halobacteri
749	5	2.6	229	1	Y755_METJA	Q57995	methanococc	822	5	2.6	247	1	ZO20_XENLA	Q9h445	halobacteri
750	5	2.6	229	1	YF67_CAEEL	Q11083	caenorhabdi	823	5	2.6	248	1	VP28_SCHPO	P18744	xenopus lae
751	5	2.6	230	1	PHOU_CAUCR	Q9ufa7	caulobacter	824	5	2.6	248	1	Y468_BORBU	O13872	schizosacch
752	5	2.6	230	1	SFSA_PYRFU	O8u1k8	pyrococcus	825	5	2.6	248	1	YDFG_ECOLI	O51424	borrelia bu
753	5	2.6	230	1	Y394_PSEAE	P24562	pseudomonas	826	5	2.6	248	1	YFGE_ECOLI	P39831	escherichia
754	5	2.6	230	1	YCAP_ECOLI	P75839	escherichia	827	5	2.6	249	1	CISH_CHICK	Q96w70	gallus gall
755	5	2.6	231	1	RHIA_RHILV	Q03313	rhizobium l	828	5	2.6	249	1	RS3_SCHPO	O60128	schizosacch
756	5	2.6	231	1	RUIC_YEAST	Q05900	saccharomyc	829	5	2.6	249	1	YAFI_ECOLI	Q47151	escherichia
757	5	2.6	231	1	SOML_SPAU	P54863	sparus aura	830	5	2.6	250	1	GLO2_BUCAP	Q08889	buchnera ap
758	5	2.6	231	1	SOM2_SPAU	P79894	sparus aura	831	5	2.6	250	1	IF2B_CAEEL	O21230	caenorhabdi
759	5	2.6	231	1	SOML_SCIOC	Q9y9k7	sciaenops o	832	5	2.6	250	1	REP2_BUCAI	Q9zey9	buchnera ap
760	5	2.6	231	1	SOML_SIGGU	Q9pwq4	siganus gut	833	5	2.6	250	1	REPA_AGRTU	P15394	agrobacteri
761	5	2.6	231	1	V054_FOWPV	P21970	fowlpox vir	834	5	2.6	251	1	YCW6_YEAST	P25659	saccharomyc
762	5	2.6	232	1	ISPD_BACSU	Q06755	bacillus su	835	5	2.6	251	1	AMPM_BORBU	O51132	borrelia bu
763	5	2.6	232	1	SPIH_HUMAN	Q99865	homo sapien	836	5	2.6	251	1	BIOC_ECOLI	P12999	escherichia

545	178	1	Y319_MYCGE	P47561 mycoplasma	618	5	2.6	202	1	DTC_RAT	P01883 rattus norv
546	178	1	YE71_METJA	Q58866 methanococc	619	5	2.6	202	1	H1L_MYTRR	Q05831 mytilus tro
547	178	1	YEVA_SCHPO	Q9p7c5 schizosacch	620	5	2.6	202	1	HAN1_CHICK	Q90691 gallus gall
548	178	1	YMF5_YEAST	Q04969 saccharomyc	621	5	2.6	202	1	IF3_MYCLE	Q9cc22 mycobacteri
549	178	1	YI29_MYCTU	Q06246 mycobacteri	622	5	2.6	202	1	VMAT_RABVN	P25224 rabies viru
550	179	1	APT_HELPJ	Q9z1g9 helicobacte	623	5	2.6	202	1	YCP2_PSEA9	P29300 pseudanabae
551	179	1	APT_HELPY	Q25296 helicobacte	624	5	2.6	202	1	YF56_MYCTU	Q10774 mycobacteri
552	179	1	DYR_MYCPU	Q98q32 mycoplasma	625	5	2.6	203	1	GSTA_RHILE	Q52828 rhizobium l
553	179	1	Y365_HAEIN	P43988 haemophilus	626	5	2.6	203	1	HM16_CAEEL	P34326 caenorhabdi
554	179	1	YNC5_CAEEL	P34538 caenorhabdi	627	5	2.6	203	1	RAS1_RHIRA	P22278 rhizomucor
555	180	1	YKG9_YEAST	P36088 saccharomyc	628	5	2.6	203	1	YAL7_METJA	Q58423 methanococc
556	182	1	KITH_AMEPV	P28852 amsacta moo	629	5	2.6	204	1	IPP2_RABIT	P11845 oryctolagus
557	182	1	RHEB_DROME	Q9vnc8 drosophila	630	5	2.6	204	1	IPP2_RAT	P50411 rattus norv
558	182	1	RRF_SYNY3	P74456 synecocyst	631	5	2.6	205	1	EF1B_YEAST	P32471 saccharomyc
559	183	1	ARL1_YEAST	P38116 saccharomyc	632	5	2.6	205	1	MD21_HUMAN	Q13257 homo sapien
560	183	1	NUSG_PASMO	Q9ck84 pasteurella	633	5	2.6	205	1	MD21_MOUSE	Q9z1b5 mus musculu
561	183	1	RL18_CICAR	Q65729 cicer ariet	634	5	2.6	205	1	RAS3_RHIRA	P22280 rhizomucor
562	183	1	YAG2_SCHPO	Q09867 schizosacch	635	5	2.6	205	1	SODM_ECO57	Q8x7b2 escherichia
563	185	1	NUSG_HAEIN	P43916 haemophilus	636	5	2.6	205	1	SODM_ECOLI	P00448 escherichia
564	185	1	RM05_ARATH	P42793 arabidopsis	637	5	2.6	205	1	SODM_SALTI	Q8z2v9 salmonella
565	185	1	RM05_BRANA	P49388 brassica na	638	5	2.6	205	1	SODM_SALTY	P43019 salmonella
566	185	1	RRF_CAMJE	Q9pif0 campylobact	639	5	2.6	205	1	YB23_METJA	Q58523 methanococc
567	186	1	ARDH_LEIDO	Q05885 leishmania	640	5	2.6	205	1	YCFC_HAEIN	P44796 haemophilus
568	186	1	ATPD_PORPU	P51243 porphyra pu	641	5	2.6	205	1	YI19_YEAST	P40500 saccharomyc
569	186	1	BPHE_COMTE	Q46373 comamonas t	642	5	2.6	207	1	FLA2_PVRHO	O58283 pyrococcus
570	186	1	INB_FELCA	Q9n210 felis silve	643	5	2.6	207	1	LEXA_BACHD	Q9ka03 bacillus ha
571	187	1	RL18_SCHPO	Q10132 schizosacch	644	5	2.6	207	1	RS8_HUMAN	P09038 homo sapien
572	187	1	TIS2_MYCPN	P75488 mycoplasma	645	5	2.6	207	1	RS8_ICTPU	Q90y96 ictalurus p
573	188	1	YEAB_SCHPO	O14076 schizosacch	646	5	2.6	207	1	SODM_CHLNP	Q929c4 chlamydia p
574	189	1	CHS1_TUBUN	P55003 tuber uncin	647	5	2.6	207	1	THIE_PVRAB	O9uzg5 pyrococcus
575	189	1	CHS2_ASPNG	P30582 aspergillus	648	5	2.6	208	1	GTH1_ARATH	P42760 arabidopsis
576	189	1	FTPA_HAEDU	O47953 haemophilus	649	5	2.6	208	1	GTP_DIRIM	P46436 dirofilaria
577	189	1	NCS1_YEAST	Q06389 saccharomyc	650	5	2.6	209	1	CHEC_BACSU	P40403 bacillus su
578	190	1	SDF_ENTHI	P34107 entamoeba h	651	5	2.6	209	1	GTP1_MOUSE	P46425 mus musculu
579	191	1	PTH_MYCTU	Q01684 saccharomyc	652	5	2.6	209	1	GTP2_MOUSE	P15157 mus musculu
580	191	1	SPR6_YEAST	Q01684 saccharomyc	653	5	2.6	209	1	GTP_BOVIN	P28801 bos taurus
581	191	1	VG27_BPMU	Q9t1w7 bacterioph	654	5	2.6	209	1	GTP_CRILLO	P46424 cricetus
582	192	1	ARL7_HUMAN	P56559 homo sapien	655	5	2.6	209	1	GTP_CRIMI	P47934 cricetus
583	192	1	Y678_AQUAE	Q66906 aquifex ao	656	5	2.6	209	1	GTP_HUMAN	P09211 homo sapien
584	193	1	BP42_PSES1	Q52439 pseudomonas	657	5	2.6	209	1	GTP_MACMU	Q28514 macaca mula
585	193	1	YF2D_CAEEL	Q09211 caenorhabdi	658	5	2.6	209	1	GTP_MESAU	Q60550 mesocricetu
586	194	1	MRP_HUMAN	P49006 homo sapien	659	5	2.6	210	1	GIDB_PASMO	P57946 pasteurella
587	194	1	YPRE_DEIRA	Q9rx68 deinococcus	660	5	2.6	210	1	OM22_ARATH	P82873 arabidopsis
588	194	1	YPBH_BACSU	P50734 bacillus su	661	5	2.6	210	1	VS10_ROTTC	P36358 porcine rot
589	195	1	PUR3_BACSU	P12040 bacillus su	662	5	2.6	211	1	SOMA_LEPOS	P79885 lepisosteus
590	195	1	WMT2_INBLE	P03493 influenza b	663	5	2.6	211	1	Y290_METJA	Q57738 methanococc
591	195	1	YOU9_CAEEL	P30636 caenorhabdi	664	5	2.6	211	1	Y290_CAEEL	P34449 caenorhabdi
592	196	1	HIS7_HALN1	Q9hn13 halobacteri	665	5	2.6	212	1	END3_RICPR	Q05956 rickettsia
593	196	1	KGUA_MYCPU	Q98pn5 mycoplasma	666	5	2.6	212	1	PCAJ_PSEPU	Q01104 pseudomonas
594	196	1	YCH6_YEAST	P25602 saccharomyc	667	5	2.6	212	1	PRL_DICLA	P48249 dicentrarch
595	197	1	GRPE_YACSK	O87776 lactobacill	668	5	2.6	212	1	SLYD_AERHY	P07046 aeromonas h
596	197	1	LEXA_THDMA	Q33927 thermotoga	669	5	2.6	212	1	SP0A_BACME	P22932 bacillus me
597	197	1	RAC1_PEA	O86948 thermotoga	670	5	2.6	212	1	TRBC_ECOLI	P18473 escherichia
598	197	1	YV26_NPVOP	Q35638 plism sativ	671	5	2.6	212	1	YQ15_CAEEL	Q19230 caenorhabdi
599	197	1	YV26_NPVOP	O10276 orgvia pseu	672	5	2.6	213	1	H1X_HUMAN	Q92532 homo sapien
600	197	1	Y979_METJA	Q58389 methanococc	673	5	2.6	213	1	HISL_LISIN	Q92e83 listeria in
601	198	1	MRP_RABIT	P25566 oryctolagus	674	5	2.6	213	1	KITH_MYCCE	P47280 mycoplasma
602	198	1	NUDH_PASMO	P57809 pasteurella	675	5	2.6	213	1	MOBE_THIFE	P22901 thiobacillu
603	198	1	OGG1_ARCFU	Q29876 archaeoglob	676	5	2.6	213	1	NQRD_CHLNP	Q9z8b4 chlamydia p
604	198	1	PRL_CHEMY	P33090 chelonias my	677	5	2.6	213	1	VE4_HPV65	Q07873 human papil
605	198	1	MRP_YEAST	P40993 saccharomyc	678	5	2.6	214	1	MAA1_CAEEL	Q18938 caenorhabdi
606	199	1	AA27_HUMAN	O60232 homo sapien	679	5	2.6	214	1	SODM_HAEIN	P43725 haemophilus
607	199	1	AA27_HUMAN	P56873 mus musculu	680	5	2.6	214	1	SODM_PASMO	Q9cpn6 pasteurella
608	199	1	MRP_MOUSE	P28667 mus musculu	681	5	2.6	215	1	EVAI_HUMAN	O60487 homo sapien
609	199	1	PTH_MYCLE	Q9cd49 mycobacteri	682	5	2.6	215	1	GLR2_ECOLI	P39811 escherichia
610	199	1	RABA_DICDI	P34141 dictyosteli	683	5	2.6	215	1	IPYR_HORVD	Q23979 hordeum vul
611	199	1	YP24_RTBVP	P27500 rice tungro	684	5	2.6	215	1	RASL_COPCI	O05058 coprinus ci
612	200	1	MAB5_CAEEL	P10038 caenorhabdi	685	5	2.6	216	1	RAS_CRYNE	O74650 cryptococcu
613	200	1	Y690_METJA	Q58102 methanococc	686	5	2.6	216	1	Y726_METJA	Q58136 methanococc
614	201	1	IF3_MYCTU	P94975 mycobacteri	687	5	2.6	217	1	MSRA_KREML	Q9apy4 corynebacte
615	201	1	RECU_LISIN	Q92ab7 listeria in	688	5	2.6	218	1	ALBR_KLEOX	P10488 klebsiella
616	201	1	YL79_YEAST	Q06252 saccharomyc	689	5	2.6	218	1	IM23_SCHJA	P27591 schistosoma
617	201	1	Z239_MOUSE	P24399 mus musculu	690	5	2.6	218	1	T2D8_HUMAN	Q12962 homo sapien



399	1	132	5	2.6	472	5	2.6	158	1	IF53_ARATH	Q9c505 arabidopsis
400	1	133	5	2.6	473	5	2.6	158	1	YJEE_HAEIN	P44492 haemophilus
401	1	133	5	2.6	474	5	2.6	159	1	FMI_MORBO	P20657 moraxella b
402	1	134	5	2.6	475	5	2.6	159	1	IF51_NICPL	P24951 nicotiana p
403	1	134	5	2.6	476	5	2.6	159	1	IF52_MEDSA	Q945f4 medicago sa
404	1	134	5	2.6	477	5	2.6	159	1	IF52_NICPL	P24922 nicotiana p
405	1	134	5	2.6	478	5	2.6	159	1	IF53_LYCES	Q9axq4 lycopersico
406	1	134	5	2.6	479	5	2.6	159	1	IF53_LYCES	Q9axq4 lycopersico
407	1	135	5	2.6	480	5	2.6	159	1	IF53_LYCES	Q9axq4 lycopersico
408	1	135	5	2.6	481	5	2.6	159	1	IF53_LYCES	Q9axq4 lycopersico
409	1	136	5	2.6	482	5	2.6	159	1	IF53_LYCES	Q9axq4 lycopersico
410	1	136	5	2.6	483	5	2.6	159	1	IF53_LYCES	Q9axq4 lycopersico
411	1	137	5	2.6	484	5	2.6	160	1	IF53_LYCES	Q9axq4 lycopersico
412	1	137	5	2.6	485	5	2.6	160	1	IF53_LYCES	Q9axq4 lycopersico
413	1	138	5	2.6	486	5	2.6	160	1	IF53_LYCES	Q9axq4 lycopersico
414	1	138	5	2.6	487	5	2.6	160	1	IF53_LYCES	Q9axq4 lycopersico
415	1	138	5	2.6	488	5	2.6	160	1	IF53_LYCES	Q9axq4 lycopersico
416	1	138	5	2.6	489	5	2.6	160	1	IF53_LYCES	Q9axq4 lycopersico
417	1	139	5	2.6	490	5	2.6	160	1	IF53_LYCES	Q9axq4 lycopersico
418	1	139	5	2.6	491	5	2.6	160	1	IF53_LYCES	Q9axq4 lycopersico
419	1	140	5	2.6	492	5	2.6	161	1	IF53_LYCES	Q9axq4 lycopersico
420	1	140	5	2.6	493	5	2.6	161	1	IF53_LYCES	Q9axq4 lycopersico
421	1	140	5	2.6	494	5	2.6	161	1	IF53_LYCES	Q9axq4 lycopersico
422	1	140	5	2.6	495	5	2.6	161	1	IF53_LYCES	Q9axq4 lycopersico
423	1	141	5	2.6	496	5	2.6	162	1	IF53_LYCES	Q9axq4 lycopersico
424	1	141	5	2.6	497	5	2.6	162	1	IF53_LYCES	Q9axq4 lycopersico
425	1	143	5	2.6	498	5	2.6	163	1	IF53_LYCES	Q9axq4 lycopersico
426	1	143	5	2.6	499	5	2.6	163	1	IF53_LYCES	Q9axq4 lycopersico
427	1	144	5	2.6	500	5	2.6	163	1	IF53_LYCES	Q9axq4 lycopersico
428	1	144	5	2.6	501	5	2.6	164	1	IF53_LYCES	Q9axq4 lycopersico
429	1	145	5	2.6	502	5	2.6	164	1	IF53_LYCES	Q9axq4 lycopersico
430	1	145	5	2.6	503	5	2.6	166	1	IF53_LYCES	Q9axq4 lycopersico
431	1	146	5	2.6	504	5	2.6	166	1	IF53_LYCES	Q9axq4 lycopersico
432	1	146	5	2.6	505	5	2.6	166	1	IF53_LYCES	Q9axq4 lycopersico
433	1	146	5	2.6	506	5	2.6	166	1	IF53_LYCES	Q9axq4 lycopersico
434	1	146	5	2.6	507	5	2.6	167	1	IF53_LYCES	Q9axq4 lycopersico
435	1	146	5	2.6	508	5	2.6	167	1	IF53_LYCES	Q9axq4 lycopersico
436	1	146	5	2.6	509	5	2.6	168	1	IF53_LYCES	Q9axq4 lycopersico
437	1	146	5	2.6	510	5	2.6	168	1	IF53_LYCES	Q9axq4 lycopersico
438	1	147	5	2.6	511	5	2.6	168	1	IF53_LYCES	Q9axq4 lycopersico
439	1	147	5	2.6	512	5	2.6	168	1	IF53_LYCES	Q9axq4 lycopersico
440	1	147	5	2.6	513	5	2.6	169	1	IF53_LYCES	Q9axq4 lycopersico
441	1	148	5	2.6	514	5	2.6	169	1	IF53_LYCES	Q9axq4 lycopersico
442	1	148	5	2.6	515	5	2.6	170	1	IF53_LYCES	Q9axq4 lycopersico
443	1	148	5	2.6	516	5	2.6	171	1	IF53_LYCES	Q9axq4 lycopersico
444	1	148	5	2.6	517	5	2.6	171	1	IF53_LYCES	Q9axq4 lycopersico
445	1	148	5	2.6	518	5	2.6	171	1	IF53_LYCES	Q9axq4 lycopersico
446	1	149	5	2.6	519	5	2.6	171	1	IF53_LYCES	Q9axq4 lycopersico
447	1	149	5	2.6	520	5	2.6	171	1	IF53_LYCES	Q9axq4 lycopersico
448	1	149	5	2.6	521	5	2.6	171	1	IF53_LYCES	Q9axq4 lycopersico
449	1	149	5	2.6	522	5	2.6	171	1	IF53_LYCES	Q9axq4 lycopersico
450	1	149	5	2.6	523	5	2.6	172	1	IF53_LYCES	Q9axq4 lycopersico
451	1	150	5	2.6	524	5	2.6	172	1	IF53_LYCES	Q9axq4 lycopersico
452	1	150	5	2.6	525	5	2.6	172	1	IF53_LYCES	Q9axq4 lycopersico
453	1	150	5	2.6	526	5	2.6	172	1	IF53_LYCES	Q9axq4 lycopersico
454	1	151	5	2.6	527	5	2.6	173	1	IF53_LYCES	Q9axq4 lycopersico
455	1	151	5	2.6	528	5	2.6	173	1	IF53_LYCES	Q9axq4 lycopersico
456	1	151	5	2.6	529	5	2.6	173	1	IF53_LYCES	Q9axq4 lycopersico
457	1	152	5	2.6	530	5	2.6	174	1	IF53_LYCES	Q9axq4 lycopersico
458	1	152	5	2.6	531	5	2.6	174	1	IF53_LYCES	Q9axq4 lycopersico
459	1	152	5	2.6	532	5	2.6	175	1	IF53_LYCES	Q9axq4 lycopersico
460	1	153	5	2.6	533	5	2.6	175	1	IF53_LYCES	Q9axq4 lycopersico
461	1	153	5	2.6	534	5	2.6	175	1	IF53_LYCES	Q9axq4 lycopersico
462	1	154	5	2.6	535	5	2.6	175	1	IF53_LYCES	Q9axq4 lycopersico
463	1	154	5	2.6	536	5	2.6	176	1	IF53_LYCES	Q9axq4 lycopersico
464	1	154	5	2.6	537	5	2.6	177	1	IF53_LYCES	Q9axq4 lycopersico
465	1	155	5	2.6	538	5	2.6	177	1	IF53_LYCES	Q9axq4 lycopersico
466	1	155	5	2.6	539	5	2.6	178	1	IF53_LYCES	Q9axq4 lycopersico
467	1	155	5	2.6	540	5	2.6	178	1	IF53_LYCES	Q9axq4 lycopersico
468	1	156	5	2.6	541	5	2.6	178	1	IF53_LYCES	Q9axq4 lycopersico
469	1	156	5	2.6	542	5	2.6	178	1	IF53_LYCES	Q9axq4 lycopersico
470	1	157	5	2.6	543	5	2.6	178	1	IF53_LYCES	Q9axq4 lycopersico
471	1	158	5	2.6	544	5	2.6	178	1	IF53_LYCES	Q9axq4 lycopersico

253	6	3.1	2774	1	MAPA_RAT	P34926	rattus norv	326	5	2.6	107	1	RP5M_AZOV1	P17160	azotobacter
254	6	3.1	2805	1	MAPA_HUMAN	P78559	homo sapien	327	5	2.6	108	1	Y095_ARCFU	Q30141	archaeoglob
255	6	3.1	3093	1	POLG_BSTV1	O65730	b genome po	328	5	2.6	108	1	YB01_SYNV3	P72745	synechocyst
256	6	3.1	3256	1	K167_HUMAN	P46013	homo sapien	329	5	2.6	108	1	YD88_SCHPO	Q10413	schizosacch
257	6	3.1	3911	1	KA99_HUMAN	Q99906	h a-kinase	330	5	2.6	109	1	CL11_PIG	Q29238	sus scrofa
258	5	2.6	15	1	UBL1_MONDO	P50103	monodelphis	331	5	2.6	109	1	KV3B_HUMAN	P01620	homo sapien
259	5	2.6	20	1	LPP2_HUMAN	P56642	homo sapien	332	5	2.6	109	1	KV3D_HUMAN	P01622	homo sapien
260	5	2.6	25	1	UBL1_BOVIN	P23356	bos taurus	333	5	2.6	109	1	KV3E_HUMAN	P01623	homo sapien
261	5	2.6	34	1	EM1_ENSM1	P27205	ensis minor	334	5	2.6	109	1	KV3G_HUMAN	P04206	homo sapien
262	5	2.6	50	1	YE24_HAEN	P44244	haemophilus	335	5	2.6	109	1	Y265_UREPA	Q99qm6	ureaplasma
263	5	2.6	54	1	RK32_TOBAC	P12198	nicotiana t	336	5	2.6	109	1	YJ49_YEAST	P47126	saccharomyc
264	5	2.6	55	1	A70A_DROME	O18666	drosophila	337	5	2.6	110	1	YG12_BPPF1	P25132	bacterioph
265	5	2.6	55	1	A70A_DROME	P05623	drosophila	338	5	2.6	110	1	YCKD_BAGSU	P42402	bacillus su
266	5	2.6	55	1	A70A_DROSE	O18417	drosophila	339	5	2.6	112	1	FMDB_METME	Q50229	methylophil
267	5	2.6	60	1	HMA2_HELTR	P17138	helobdella	340	5	2.6	114	1	MYH8_MOUSE	P13542	mus muscucu
268	5	2.6	62	1	RL28_LISMO	Q92a12	listeria mo	341	5	2.6	114	1	RL19_BACHD	Q9ka16	bacillus ha
269	5	2.6	62	1	GBAK_CRIGR	Q60397	cricetulus	342	5	2.6	116	1	RL31_CHLRE	P45841	chlamydomon
270	5	2.6	64	1	RK35_CYACA	Q9tir9	cyandium c	343	5	2.6	116	1	RR8_MUSAC	Q42362	musa acumin
271	5	2.6	70	1	Y40N_RHISN	P55599	rhizobium s	344	5	2.6	116	1	YC20_MESVI	Q9mul5	mesostigma
272	5	2.6	72	1	RL15_BACLI	P35138	bacillus li	345	5	2.6	116	1	YG10_YEAST	P53222	saccharomyc
273	5	2.6	73	1	RL15_STACA	P35139	staphylococ	346	5	2.6	116	1	YJG7_YEAST	P40365	saccharomyc
274	5	2.6	76	1	SECG_BACSU	Q32233	bacillus su	347	5	2.6	118	1	YJG7_YEAST	Q09774	schizosacch
275	5	2.6	76	1	UBIQ_ACACA	P49634	acanthamoeb	348	5	2.6	119	1	ACPS_STRAA	Q99s14	staphylococ
276	5	2.6	76	1	UBIQ_ACECL	P42739	acetabulari	349	5	2.6	119	1	ACPS_STAAU	Q9zah6	staphylococ
277	5	2.6	76	1	UBIQ_AGLNE	P42740	aglaothami	350	5	2.6	120	1	YF14_MYCPN	P75272	mycoplasma
278	5	2.6	76	1	UBIQ_CABEL	P14792	caenorhabdi	351	5	2.6	120	1	YQ58_MYCTU	P71955	mycobacteri
279	5	2.6	76	1	UBIQ_CHLRE	P14624	chlamydomon	352	5	2.6	121	1	RL14_AQUAE	O67570	aquifex aeo
280	5	2.6	76	1	UBIQ_COPCO	P19848	coprinus co	353	5	2.6	121	1	RL14_AOUPY	Q92i42	aquifex pyr
281	5	2.6	76	1	UBIQ_DICDI	P08618	dictyosteli	354	5	2.6	121	1	VNBP_POPMV	Q02123	poplar mosa
282	5	2.6	76	1	UBIQ_DROME	Q9vz14	drosophila	355	5	2.6	121	1	YIF9_YEAST	P40520	saccharomyc
283	5	2.6	76	1	UBIQ_EIMBO	P46574	eimeria bov	356	5	2.6	121	1	YIF9_YEAST	P06381	marchantia
284	5	2.6	76	1	UBIQ_EUPEU	P23324	euplotes eu	357	5	2.6	122	1	RL14_MYCTU	P95068	mycobacteri
285	5	2.6	76	1	UBIQ_HUMAN	P02248	homo sapien	358	5	2.6	122	1	YV66_METTF	P29580	methanobact
286	5	2.6	76	1	UBIQ_LEIMA	Q05550	leishmania	359	5	2.6	122	1	YV66_METTF	P29581	methanobact
287	5	2.6	76	1	UBIQ_LEITA	P49635	leishmania	360	5	2.6	123	1	RS12_CHLPN	Q92800	chlamydia p
288	5	2.6	76	1	UBIQ_NEUCR	P13117	neurospora	361	5	2.6	123	1	UL59_RCMVA	P16746	human cytom
289	5	2.6	76	1	UBIQ_PHYIN	P22589	phytophthor	362	5	2.6	124	1	SOR_PYRFU	P82385	pyrococcus
290	5	2.6	76	1	UBIQ_SOYBN	P03993	glycine max	363	5	2.6	124	1	SPH2_SULSO	Q9uwy8	sulfolobus
291	5	2.6	76	1	UBIQ_STRPU	P23398	strongyloce	364	5	2.6	124	1	YC33_CHLRE	P05724	chlamydomon
292	5	2.6	76	1	UBIQ_TETPY	P20685	tetrahymena	365	5	2.6	125	1	ALR_MOUSE	P56213	mus muscucu
293	5	2.6	76	1	UBIQ_TRYBB	P15174	trypanosoma	366	5	2.6	125	1	ALR_MOUSE	Q93042	rattus norv
294	5	2.6	76	1	UBIQ_TRYCR	P08565	trypanosoma	367	5	2.6	125	1	H2BH_HUMAN	Q93078	homo sapien
295	5	2.6	76	1	UBIQ_YEAST	P04838	saccharomyc	368	5	2.6	125	1	PYP_CHRSA	P81046	chromatium
296	5	2.6	77	1	UBIL_NPVAC	P16709	autographa	369	5	2.6	125	1	RPAG_YEAST	P32529	saccharomyc
297	5	2.6	82	1	Y235_METJA	Q57687	methanococ	370	5	2.6	125	1	Y364_AQUAE	O56689	aquifex aeo
298	5	2.6	83	1	PYS1_SYNV3	P73202	synechocyst	371	5	2.6	125	1	YCX1_GUTH	O78420	guillardia
299	5	2.6	84	1	ACPH_MYCGE	P47529	mycoplasma	372	5	2.6	125	1	YEGR_ECOLI	P76406	escherichia
300	5	2.6	84	1	YCX9_OENHO	Q9mtm6	oenothera h	373	5	2.6	126	1	PAND_YERPE	Q8zbn6	yersinia pe
301	5	2.6	87	1	VAPL_BP186	P21681	bacterioph	374	5	2.6	126	1	YK50_ARCFU	O28229	archaeoglob
302	5	2.6	87	1	VE4_HPV51	P26548	human papil	375	5	2.6	127	1	RS8E_PYRHO	O58830	pyrococcus
303	5	2.6	90	1	RPOL_SULAC	P46217	sulfolobus	376	5	2.6	127	1	RT08_ACACA	P46757	acanthamoeb
304	5	2.6	91	1	SR19_METTH	O26267	methanobact	377	5	2.6	127	1	RT12_CHOCR	P48858	chondrus cr
305	5	2.6	92	1	RR19_MALZE	P06588	zea mays (m	378	5	2.6	127	1	TRM1_ECOLI	P10026	escherichia
306	5	2.6	92	1	RR19_ORYSA	P12153	oryza sativ	379	5	2.6	127	1	TRM6_ECOLI	P18808	escherichia
307	5	2.6	92	1	RR19_SECEC	P33954	secale cere	380	5	2.6	127	1	TRM7_ECOLI	P33787	escherichia
308	5	2.6	93	1	SP3D_BACSU	P15281	bacillus su	381	5	2.6	127	1	TRM9_ECOLI	P18807	escherichia
309	5	2.6	93	1	UBIL_NPVOP	Q05120	orgyia pseu	382	5	2.6	128	1	KV3K_HUMAN	P06311	homo sapien
310	5	2.6	96	1	VG26_BPMU	Q9t1w8	bacterioph	383	5	2.6	128	1	Y097_BORBU	O51124	borrelia bu
311	5	2.6	96	1	VPR_SIVCZ	P17287	chimpanzee	384	5	2.6	128	1	Y358_METJA	Q37804	methanococc
312	5	2.6	96	1	Y448_AQUAE	O66756	aquifex aeo	385	5	2.6	128	1	Y5802_METJA	Q58002	methanococc
313	5	2.6	96	1	YEBG_ECOLI	P33220	escherichia	386	5	2.6	128	1	Y040_BACAN	Q9rmz2	bacillus su
314	5	2.6	97	1	CLU83_HSV62	P52461	human herpe	387	5	2.6	129	1	PLGB_BAGSU	Q20510	bacillus su
315	5	2.6	97	1	RL28_BRUME	Q8vjme	brucella me	388	5	2.6	129	1	GSPL_PSPAE	Q00516	pseudomonas
316	5	2.6	98	1	ACYO_CHICK	P07032	gallus gall	389	5	2.6	129	1	H3_LEIIN	P40285	leishmania
317	5	2.6	98	1	NULM_HIFAM	Q92zy3	hippopotamu	390	5	2.6	129	1	HYTA_APIWE	Q10416	apis mellif
318	5	2.6	98	1	VG04_BPMD2	O64200	mycobacteri	391	5	2.6	129	1	KV3L_HUMAN	P18135	homo sapien
319	5	2.6	99	1	ILB6_CABEL	P56173	caenorhabdi	392	5	2.6	129	1	KV3M_HUMAN	P18136	homo sapien
320	5	2.6	99	1	PSD2_BOVIN	P56701	bos taurus	393	5	2.6	129	1	YG2Z_YEAST	P53268	saccharomyc
321	5	2.6	100	1	KV3C_HUMAN	P01621	homo sapien	394	5	2.6	130	1	YQ26_BACAN	Q9rn06	bacillus an
322	5	2.6	102	1	YEG5_YEAST	P43539	saccharomyc	395	5	2.6	131	1	RSBU_SYNV3	Q55332	synechocyst
323	5	2.6	105	1	NDCL_BRAJA	P6024	bradyrhizob	396	5	2.6	131	1	RS12_DEIRA	Q9rxk7	deinococcus
324	5	2.6	107	1	RS20_THETH	P80380	thermus the	397	5	2.6	131	1	YC85_METTH	P72346	methanobact
325	5	2.6	107	1	E312_ADE05	P06496	human adeno	398	5	2.6	132	1	RS8_BORBU	O51445	borrelia bu

107	1	402	3..1	6	180	6	3..1	690	1	PCD8_DROME	Q9vq79 drosophila
108	1	403	3..1	6	181	6	3..1	692	1	YFB3_YEAST	P43579 saccharomyc
109	6	406	3..1	6	182	6	3..1	710	1	Y026_YEAST	P39744 saccharomyc
110	1	407	3..1	6	183	6	3..1	717	1	MCCA_MOUSE	Q99nr8 mus musculus
111	6	425	3..1	6	184	6	3..1	727	1	CTCF_HUMAN	P49711 homo sapien
112	6	425	3..1	6	185	6	3..1	728	1	CTCF_CHICK	Q08705 gallus gall
113	6	425	3..1	6	186	6	3..1	733	1	K6A2_MOUSE	Q9wtu3 mus musculus
114	6	425	3..1	6	187	6	3..1	752	1	PHSM_STRPN	P29849 streptococc
115	6	426	3..1	6	188	6	3..1	768	1	CNIC_RAT	Q63421 rattus norv
116	6	431	3..1	6	189	6	3..1	790	1	CB80_HUMAN	Q09161 homo sapien
117	6	439	3..1	6	190	6	3..1	805	1	CC48_CAPAN	Q96372 capsicum an
118	6	441	3..1	6	191	6	3..1	806	1	MK07_MOUSE	Q9wvs8 mus musculus
119	6	441	3..1	6	192	6	3..1	807	1	CC48_SOYBN	Q47774 glycine max
120	6	443	3..1	6	193	6	3..1	809	1	CC48_ARATH	P54609 arabidopsis
121	6	444	3..1	6	194	6	3..1	809	1	HRPB_ECOLI	P37024 escherichia
122	6	449	3..1	6	195	6	3..1	815	1	MY07_HUMAN	Q13164 homo sapien
123	6	449	3..1	6	196	6	3..1	827	1	GYRA_HELPY	P48370 helicobacte
124	6	457	3..1	6	197	6	3..1	834	1	PLSB_PSEAE	Q9hbw7 pseudomonas
125	6	461	3..1	6	198	6	3..1	835	1	CC48_YEAST	P25694 saccharomyc
126	6	463	3..1	6	199	6	3..1	844	1	PHSG_DROME	Q9xt19 drosophila
127	6	465	3..1	6	200	6	3..1	855	1	NOL1_HUMAN	P46087 homo sapien
128	6	467	3..1	6	201	6	3..1	858	1	CNRA_BOVIN	P11541 bos taurus
129	6	471	3..1	6	202	6	3..1	858	1	CNRA_MOUSE	P27664 mus musculus
130	6	475	3..1	6	203	6	3..1	859	1	CNRA_HUMAN	P16499 homo sapien
131	6	475	3..1	6	204	6	3..1	860	1	CNRA_CANFA	Q28463 canis famli
132	6	476	3..1	6	205	6	3..1	863	1	APCE_GALSU	P35911 galleria s
133	6	478	3..1	6	206	6	3..1	864	1	YG3M_YEAST	P48237 saccharomyc
134	6	481	3..1	6	207	6	3..1	875	1	HELX_SULSO	P95949 sulfolobus
135	6	484	3..1	6	208	6	3..1	882	1	DPO1_SULSO	P26811 sulfolobus
136	6	484	3..1	6	209	6	3..1	899	1	YMJ3_YEAST	Q04500 saccharomyc
137	6	485	3..1	6	210	6	3..1	906	1	FOX2_CANTR	P22414 candida tro
138	6	489	3..1	6	211	6	3..1	924	1	DPOL_BPSPI	P30314 bacterioph
139	6	494	3..1	6	212	6	3..1	928	1	RR44_HUMAN	Q9y211 homo sapien
140	6	496	3..1	6	213	6	3..1	960	1	DLG1_DROME	P31007 drosophila
141	6	501	3..1	6	214	6	3..1	994	1	DPO2_KLULA	P05468 kluyveromyc
142	6	501	3..1	6	215	6	3..1	994	1	SVV_XLFLA	Q9ph12 xylella fas
143	6	501	3..1	6	216	6	3..1	1011	1	VPS3_YEAST	P23643 saccharomyc
144	6	503	3..1	6	217	6	3..1	1021	1	MAFA_MOUSE	Q9qyr6 mus musculus
145	6	506	3..1	6	218	6	3..1	1053	1	SPS_SOLTU	Q43845 solanum tub
146	6	508	3..1	6	219	6	3..1	1057	1	ANPA_MOUSE	P18293 mus musculus
147	6	514	3..1	6	220	6	3..1	1057	1	ANPA_RAT	P18910 rattus norv
148	6	514	3..1	6	221	6	3..1	1061	1	ANPA_HUMAN	P16066 homo sapien
149	6	515	3..1	6	222	6	3..1	1084	1	XP01_YEAST	P30822 saccharomyc
150	6	516	3..1	6	223	6	3..1	1091	1	JSN1_YEAST	P47135 saccharomyc
151	6	519	3..1	6	224	6	3..1	1102	1	MYSC_CHICK	P29616 gallus gall
152	6	525	3..1	6	225	6	3..1	1113	1	N116_YEAST	Q02630 saccharomyc
153	6	529	3..1	6	226	6	3..1	1119	1	RPOB_THEAQ	Q9kwu7 thermus aqu
154	6	530	3..1	6	227	6	3..1	1176	1	HMDH_PHYBL	Q12649 phycomyces
155	6	548	3..1	6	228	6	3..1	1210	1	AF4_HUMAN	P51825 homo sapien
156	6	549	3..1	6	229	6	3..1	1220	1	ATB3_HUMAN	Q16720 homo sapien
157	6	549	3..1	6	230	6	3..1	1258	1	ATB3_RAT	Q64568 rattus norv
158	6	549	3..1	6	231	6	3..1	1335	1	DP3A_DEIRA	Q9rx08 deinococcus
159	6	553	3..1	6	232	6	3..1	1345	1	IF3A_MOUSE	P23116 mus musculus
160	6	555	3..1	6	233	6	3..1	1403	1	YGNI_YEAST	Q91293 rana catesb
161	6	556	3..1	6	234	6	3..1	1407	1	TRHY_RABIT	P37709 saccharomyc
162	6	565	3..1	6	235	6	3..1	1476	1	AT7A_CRIGR	P49015 cricetus
163	6	568	3..1	6	236	6	3..1	1477	1	YORI_YEAST	P53049 saccharomyc
164	6	569	3..1	6	237	6	3..1	1489	1	YGPI_YEAST	P53115 saccharomyc
165	6	576	3..1	6	238	6	3..1	1491	1	AT7A_MOUSE	Q64430 mus musculus
166	6	577	3..1	6	239	6	3..1	1496	1	CPSM_RANCA	Q91293 ovis aries
167	6	599	3..1	6	240	6	3..1	1549	1	TRHY_SHEEP	P22793 saccharomyc
168	6	600	3..1	6	241	6	3..1	1564	1	PDRA_YEAST	P51533 saccharomyc
169	6	603	3..1	6	242	6	3..1	1653	1	CLH_YEAST	P22137 saccharomyc
170	6	620	3..1	6	243	6	3..1	1850	1	VIT2_CHICK	P02845 gallus gall
171	6	620	3..1	6	244	6	3..1	1898	1	TRHY_HUMAN	Q07283 homo sapien
172	6	620	3..1	6	245	6	3..1	1906	1	YFAQ_ANASP	Q8ym40 anabaena sp
173	6	621	3..1	6	246	6	3..1	1939	1	MYH6_MESAU	P13539 mesocricetu
174	6	622	3..1	6	247	6	3..1	2052	1	FYVI_MOUSE	Q921t6 mus musculus
175	6	639	3..1	6	248	6	3..1	2349	1	TPR_HUMAN	P12270 homo sapien
176	6	642	3..1	6	249	6	3..1	2363	1	SPCO_MOUSE	Q62261 mus musculus
177	6	645	3..1	6	250	6	3..1	2364	1	SPCO_HUMAN	Q01082 homo sapien
178	6	659	3..1	6	251	6	3..1	2388	1	SPCP_RAT	Q9qwn8 rattus norv
179	6	689	3..1	6	252	6	3..1	2390	1	SPCP_HUMAN	Q15020 homo sapien

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:55:13 ; Search time 13 seconds  
(without alignments)

612.573 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 192

Sequence: 1 MRKEVPEMLNKNYKPGPQF.....EEODKEMTSKQHLFVRKN 192

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	4.2	733	1 K6A2_HUMAN	Q15349 homo sapien
2	7	3.6	57	1 RL32_MYCTU	P58287 mycobacteri
3	7	3.6	81	1 SFR6_RABIT	O18776 oryctolagus
4	7	3.6	124	1 RL12_PORPU	P51289 porphyra pu
5	7	3.6	186	1 TATB_HABIN	P57047 haemophilus
6	7	3.6	200	1 RACH_DICDI	Q59pr7 dictyosteli
7	7	3.6	344	1 SFR6_HUMAN	Q13247 homo sapien
8	7	3.6	381	1 DHPH_BACSH	P23307 bacillus sp
9	7	3.6	382	1 NCAP_CVPPF	P05991 porcine tra
10	7	3.6	382	1 NCAP_CVPPU	P04134 porcine tra
11	7	3.6	382	1 NCAP_CVPR8	P33463 porcine res
12	7	3.6	382	1 NCAP_CVPRM	P24411 porcine res
13	7	3.6	577	1 YX8_YEAST	P53075 saccharomyc
14	7	3.6	632	1 PTMA_BUCAI	P57635 buchnera ap
15	7	3.6	788	1 KDGG_MOUSE	Q91wg7 mus musculu
16	7	3.6	788	1 KDGG_RAT	P49620 rattus norv
17	7	3.6	791	1 KDGG_HUMAN	P49619 homo sapien
18	7	3.6	798	1 L100_ADEGX	P36856 avian adeno
19	7	3.6	859	1 STL_BUCAI	P57519 buchnera ap
20	7	3.6	1233	1 IRS1_MOUSE	P35569 mus musculu
21	7	3.6	1242	1 IRS1_HUMAN	P35568 homo sapien
22	7	3.6	1318	1 VIVD_BPT7	P03726 bacterioph
23	7	3.6	1321	1 IRS2_MOUSE	P81122 mus musculu
24	7	3.6	1324	1 IRS2_HUMAN	Q9y4h2 homo sapien
25	7	3.6	2145	1 U520_CABEL	Q02920 caenorhabdi
26	7	3.6	2386	1 RAD3_SCHPO	Q02099 schizosacch
27	6	3.1	35	1 PH11_MYCTU	P35422 mytilus cal
28	6	3.1	64	1 YW83_MYCTU	Q50682 mycobacteri
29	6	3.1	75	1 R14E_METH	O26139 methanobact
30	6	3.1	91	1 PH11_MYTED	Q04621 mytilus edu
31	6	3.1	103	1 CYAY_RICCN	Q921h7 rickettsia
32	6	3.1	118	1 Y781_SULSO	Q9uxg8 sulfolobus
33	6	3.1	120	1 YIBI_ECOLI	P32108 escherichia

34	3.1	122	1 YLB9_CABEL	P46583 caenorhabdi
35	3.1	125	1 H2B1_XENLA	P02281 xenopus lae
36	3.1	125	1 H2B2_XENLA	P06900 xenopus lae
37	3.1	126	1 GP48_BPSF1	O48402 bacterioph
38	3.1	126	1 RK12_CYAPA	P48124 cyanophora
39	3.1	148	1 NDK_THEAC	Q9hj59 thermoplas
40	3.1	149	1 RL13_CHLPM	Q928t7 chlamydia p
41	3.1	150	1 RL13_CHLPM	Q928t7 chlamydia p
42	3.1	150	1 RL13_CHLTR	O84127 chlamydia t
43	3.1	150	1 YDPC_SCHPO	O14015 schizosacch
44	3.1	155	1 REG1_PYRAB	Q9v2m0 pyrococcus
45	3.1	155	1 REG1_PYRHO	P55685 pyrococcus
46	3.1	181	1 Y4WG_RHISN	P55685 rhizobium s
47	3.1	187	1 RK24_TOBAC	Q02764 nicotiana t
48	3.1	188	1 NRFX_PASMU	Q9cpk9 pasteurella
49	3.1	191	1 Y948_METJA	Q58358 methanococc
50	3.1	194	1 RS7_FUGRU	P50894 fugu rubrip
51	3.1	202	1 RR4_MARPO	P06358 marchantia
52	3.1	205	1 RR4_EUGGR	P27418 euglena gra
53	3.1	206	1 Y140_METJA	Q57605 methanococc
54	3.1	210	1 GTS1_CABEL	Q09607 caenorhabdi
55	3.1	210	1 THIE_ARCFU	Q28205 archaeoglob
56	3.1	215	1 ERD2_PETHY	Q92tn2 petunia hyb
57	3.1	221	1 YR75_CABEL	Q09394 caenorhabdi
58	3.1	232	1 SLT2_CABEL	Q19892 caenorhabdi
59	3.1	249	1 KTXG_KIULA	P09807 kluyveromyc
60	3.1	259	1 Y068_CHLTR	O84071 chlamydia t
61	3.1	261	1 YEV6_YEAST	P40078 saccharomyc
62	3.1	271	1 CEAM_ECOLI	P05820 escherichia
63	3.1	280	1 LBX1_HUMAN	P52954 homo sapien
64	3.1	280	1 LBX1_MOUSE	P52955 mus musculu
65	3.1	280	1 LGTE_NEIMB	Q51117 neisseria m
66	3.1	280	1 YB27_METJA	O58527 methanococc
67	3.1	289	1 GNPI_HUMAN	P46926 homo sapien
68	3.1	292	1 YQ14_CABEL	Q09505 caenorhabdi
69	3.1	294	1 DEH1_MORSP	Q01398 moraxella s
70	3.1	294	1 YQ19_CABEL	Q09507 caenorhabdi
71	3.1	301	1 ORT2_HUMAN	Q9bxi2 homo sapien
72	3.1	308	1 JUNB_CYPCA	P79703 cyprinus ca
73	3.1	312	1 O1F1_HUMAN	O43749 homo sapien
74	3.1	314	1 ARCC_LACSK	O53090 lactobacill
75	3.1	323	1 JUND_CHICK	P27921 gallus gall
76	3.1	326	1 RNHL_SCHPO	Q10236 schizosacch
77	3.1	327	1 E13D_HORVU	Q02437 hordeum vul
78	3.1	330	1 YN7_YEAST	P50947 saccharomyc
79	3.1	331	1 YZ44_METJA	O50299 methanococc
80	3.1	332	1 BIOB_BACSH	P19206 bacillus sp
81	3.1	341	1 JUND_MOUSE	P15066 mus musculu
82	3.1	341	1 JUND_RAT	P52909 rattus norv
83	3.1	344	1 Y182_ARCFU	O30055 archaeoglob
84	3.1	345	1 PRIM_PYRAB	Q9v292 pyrococcus
85	3.1	347	1 JUND_HUMAN	P17535 homo sapien
86	3.1	353	1 GB11_CAVPO	P38401 cavia porce
87	3.1	353	1 GB11_CHICK	P50146 gallus gall
88	3.1	353	1 GB11_HUMAN	P04898 homo sapien
89	3.1	353	1 GB11_ORYLA	P87383 oryzaes lat
90	3.1	353	1 GB11_RAT	P10824 rattus norv
91	3.1	353	1 GB11_XENLA	P27044 xenopus lae
92	3.1	353	1 PLSA_AGR5	Q8ug63 agrobacteri
93	3.1	354	1 GB12_CANFA	P38400 canis fami
94	3.1	354	1 GB12_CAVPO	P38402 cavia porce
95	3.1	354	1 GB12_CHICK	P50147 gallus gall
96	3.1	354	1 GB12_HUMAN	P04899 homo sapien
97	3.1	354	1 GB12_MOUSE	P08752 mus musculu
98	3.1	354	1 GB12_RAT	P04897 rattus norv
99	3.1	356	1 QUEA_RALSO	Q8xv55 ralstonia s
100	3.1	360	1 FWLL_BACLI	P36550 bacillus li
101	3.1	360	1 FLHB_BACSU	P35538 bacillus su
102	3.1	370	1 CYSA_MARPO	P10091 marchantia
103	3.1	373	1 DP3B_MYCPU	Q58r66 mycoplasma
104	3.1	381	1 YNU6_CABEL	P50444 caenorhabdi
105	3.1	391	1 IF43_NICPL	P41380 nicotiana p
106	3.1	393	1 VN53_ROTPO	Q00033 porcine rot

THIS PAGE BLANK (USPTO)

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSOSS 138  
|||||||  
Db 267 RSKSOSS 273

RESULT 40  
AAR67708  
ID AAR67708 standard; Protein; 1243 AA.  
XX  
AC AAR67708;  
XX  
DT 20-JUL-1995 (first entry)  
XX  
DE Insulin receptor substrate-1.  
XX  
KW Insulin receptor substrate-1; IRS-1;  
KW non-insulin-dependent diabetes; NIDDM.  
XX  
OS Homo sapiens.  
XX  
FN WO9429345-A.  
XX  
PD 22-DEC-1994.  
XX  
PF 10-JUN-1994; 94WO-DK00227.  
XX  
PR 10-JUN-1993; 93DK-0000683.  
PR 09-AUG-1993; 93DK-0000915.  
XX  
PA (NOVO ) NOVO-NORDISK AS.  
XX  
PI Bjoerbaek C, Frederiksen KA, Pedersen O;  
XX  
DR WPI: 1995-036405/05.  
DR N-PSDB; AAQ75977.  
XX  
PT DNA encoding mutated insulin receptor substrate 1 - and related  
PT vectors, transformed cells and mutant proteins, useful for  
PT detecting predisposition to non-insulin-dependent diabetes  
XX  
PS Disclosure; Page 31-38; 58pp; English.  
XX  
CC Genomic DNA from human leukocyte nuclei was subjected to PCR  
CC amplification using primers (given in AAQ75978-Q76015) specific for the  
CC human IRS-1 gene (AAQ75977). Amplified DNA was subjected to SSCP  
CC analysis to detect mutations in the IRS-1 gene, indicating a  
CC predisposition to NIDDM.  
XX  
SQ Sequence 1243 AA;

Query Match 3.6%; Score 7; DB 16; Length 1243;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSOSS 138  
|||||||  
Db 268 RSKSOSS 274

Search completed: February 26, 2003, 10:00:07  
Job time : 66 secs

PS Claim 4; Fig 16; 86pp; English.

CC The invention relates to modulators that affect the interaction of a  
 CC polypeptide defining a PI (phosphotyrosine interaction) domain, or its  
 CC conserved variants or fragments, with APP (amyloid precursor protein).  
 CC The polypeptide defining a PI domain binds to the cytoplasmic domain of  
 CC APP, regulating the secretion of APP fragments. By modulating this  
 CC interaction, the modulator can inhibit or prevent APP processing and  
 CC trafficking, and thus progress or onset of Alzheimer's disease. The  
 CC modulators are used to treat and/or prevent neurodegenerative diseases  
 CC in mammals, particularly Alzheimer's disease and dementia. Sequences  
 CC AAV13448-461 represent examples of polypeptides defining a PI domain.

XX  
 SQ Sequence 1242 AA;

Query Match 3.6%; Score 7; DB 20; Length 1242;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138  
 | | | | | | |  
 Db 267 RSKSQSS 273

RESULT 38

AAW93972  
 ID AAW93972 standard; protein; 1242 AA.

AC AAW93972;

XX  
 XX  
 XX 30-JUN-1999 (first entry)

DE Human IRS-1 and IRS-2 binding inhibitor protein.

XX  
 XX Insulin receptor substrate; IRS-1; IRS-2; inhibitor; disease; obesity;  
 KW insulin resistance; 14-3-3 protein; inhibitor; screening; diabetes;  
 KW diabetic retinopathy; diabetic neuropathy; glucose tolerance;  
 KW diabetic nephropathy; hyperinsulinaemia; hyperlipemia; arteriosclerosis;  
 KW hypertension; ischaemic heart disease; ischaemic brain disease;  
 KW peripheral embolism.

XX  
 OS Homo sapiens.

XX  
 XX WO9916462-A1.

PN  
 XX  
 PD 08-APR-1999.

XX  
 XX 25-SEP-1998; 98WO-JP04293.

PF  
 XX 29-SEP-1997; 97JP-0263719.

PR  
 XX (DAUC ) DAIICHI PHARM CO LTD.

PA  
 XX Asano T, Kanda A, Kubo H, Yazaki Y;  
 PI WPI; 1999-254929/21.

XX  
 XX Treatment of insulin resistance using insulin receptor substrate  
 PT binding inhibitor

XX  
 XX Disclosure; Page 16-22; 30pp; Japanese.

PS  
 CC This invention describes a method for the treatment of diseases involving  
 CC insulin resistance using a substance which inhibits the binding of  
 CC insulin receptor substrate to 14-3-3 protein, identified by screening  
 CC potential inhibitors for their ability to inhibit this binding. The  
 CC composition described in the invention for the treatment of diseases  
 CC involving insulin resistance contains as active component, an inhibitor  
 CC of the binding of all or part of insulin receptor substrate 1 or 2  
 CC (IRS-1, IRS-2) to all or part of 14-3-3 protein. Disorders involving  
 CC insulin resistance include diabetes, diabetic retinopathy, diabetic  
 CC neuropathy, impaired glucose tolerance, diabetic nephropathy,  
 CC hyperinsulinaemia, hyperlipemia, arteriosclerosis, hypertension,

CC obesity, ischaemic heart disease, ischaemic brain disease and peripheral  
 CC embolism.

XX  
 SQ Sequence 1242 AA;

Query Match 3.6%; Score 7; DB 20; Length 1242;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138  
 | | | | | | |  
 Db 267 RSKSQSS 273

RESULT 39

AAB83921  
 ID AAB83921 standard; Protein; 1242 AA.

XX  
 XX AAB83921;

AC  
 XX  
 XX 23-JUL-2001 (first entry)

DT  
 XX Amino acid sequence of a human insulin receptor substrate (IRS).

DE  
 XX Epidermal growth factor-like domain; EGF-like domain; cancer;  
 KW human aspartyl beta-hydroxylase; HAAH; malignant neoplasm; tumour;  
 KW insulin receptor substrate; IRS.

XX  
 OS Homo sapiens.

XX  
 XX WO200135102-A2.

PN  
 XX  
 PD 17-MAY-2001.

XX  
 XX 08-NOV-2000; 2000WO-US30738.

PF  
 XX 08-NOV-1999; 99US-0436184.

PR  
 XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

PA  
 XX Wands JR, De La Monte SM, Ince N, Carlson RI;  
 PI WPI; 2001-329171/34.

DR  
 XX N-PSDB; AAF89816.

XX  
 XX Diagnosing malignant neoplasm in a mammal, involves contacting  
 PT mammalian sample with antibody that binds to human aspartyl  
 PT beta-hydroxylase polypeptide to form antigen-antibody complex and  
 PT detecting the complex

XX  
 XX Example 7; Page 52; 76pp; English.

PS  
 XX The present sequence represents a human insulin receptor substrate (IRS).  
 CC Compounds which inhibit IRS also inhibit growth of tumours which  
 CC overexpress human aspartyl (asparaginyl) beta-hydroxylase (HAAH) enzyme.  
 CC Epidermal growth factor (EGF)-like domains of polypeptides are  
 CC hydroxylated by HAAH enzymes. HAAH is used in the method of the  
 CC invention. The specification describes a method for diagnosing a  
 CC malignant neoplasm in a mammal. The method comprises contacting a body  
 CC fluid with an antibody which binds to HAAH polypeptide under complex  
 CC forming conditions, and detecting the antigen-antibody complex. The  
 CC method is useful for diagnosing and prognosing a malignant neoplasm in  
 CC a bodily fluid e.g. central nervous system (CNS)-derived body fluid,  
 CC blood, serum, urine, saliva, sputum, lung effusion, and ascites fluid  
 CC of mammal, where the neoplasm is derived from endodermal tissue and is  
 CC selected from colon cancer, breast cancer, pancreatic cancer, liver  
 CC cancer, cancer of bile ducts, and cancer of the CNS. HAAH antibodies,  
 CC linked to a cytotoxic agent, are useful for killing tumour cells.

XX  
 SQ Sequence 1242 AA;

Query Match 3.6%; Score 7; DB 22; Length 1242;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;

PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160985.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 3.6%; Score 7; DB 21; Length 823;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AKQHLUF 188  
 Db 27 AKQHLUF 33

RESULT 36  
 AAB42816  
 ID AAB42816 standard; Protein; 849 AA.

XX AAB42816;  
 XX  
 DT 08-FEB-2001 (first entry)

DE Human ORFX ORF2580 polypeptide sequence SEQ ID NO:5160.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antiposioratic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antirheumatic;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

XX Homo sapiens.  
 XX WO200058473-A2.  
 XX  
 XX 05-OCT-2000.  
 XX  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX  
 XX 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.  
 XX Shimkets RA, Leach M;  
 PI WPI; 2000-602362/57.  
 DR N-PSDB; AAC77025.  
 XX

PT Novel nucleic acids and peptides derived from open reading frame X,  
 useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX Claim 11; Page 4342-4344; 5507pp; English.

XX AAC77446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiposioratic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 849 AA;

Query Match 3.6%; Score 7; DB 21; Length 849;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 DWGNEQL 72  
 Db 33 DWGNEQL 39

RESULT 37  
 AAY13461  
 ID AAY13461 standard; protein; 1242 AA.

XX AAY13461;  
 XX 26-JUL-1999 (first entry)  
 XX Amino acid sequence of human IRS-1 protein.  
 DE Phosphotyrosine interaction domain; PI domain; modulator; APP;  
 XX amyloid precursor protein; cytoplasmic domain; Alzheimer's disease;  
 KW neurodegenerative disease; dementia; IRS-1.

XX Homo sapiens.  
 XX WO9921995-A1.  
 XX 06-MAY-1999.  
 XX 23-OCT-1998; 98WO-US22523.  
 XX 17-APR-1998; 98US-0062085.  
 PR 24-OCT-1997; 97US-0957660.  
 XX (UYRQ ) UNIV ROCKEFELLER.

XX Buxbaum J, Greengard P, Sabo S;  
 PI WPI; 1999-326703/27.  
 DR Cell line for screening agents that inhibit processing of amyloid  
 PT precursor protein  
 XX



PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139751.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0132407.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 15-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149802.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

XX Claim 6; SEQ ID No 1353; 2504pp; French.  
 PS The present invention is related to a *Lactococcus lactis* nucleotide  
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC WO200177334 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 628 AA;  
 Query Match 3.6%; Score 7; DB 23; Length 628;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 53 FSEILLK 59  
 DB 413 FSEILLK 419  
 RESULT 34  
 AAU81501  
 ID AAU81501 standard; Protein; 782 AA.  
 XX  
 AC AAU81501;  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Chlamydomonas interflagellar transport protein, IFT88.  
 KW Human; chlamydomonas; intraflagellar transport protein; Che-2;  
 KW antibacterial; protozoa; contraceptive; antiinflammatory; malaria;  
 KW Chlamydomonas; IFT; mouse; kidney disease; retinal disorder; nematode;  
 KW thyroid disorder; chondrocyte disease; olfactory disease; azoospermia;  
 KW primary ciliary dyskinesia; insect; protozoa; male contraceptive;  
 KW parasitic disease; African sleeping sickness; trypanosomiasis;  
 KW leishmaniasis; trichomonosias; giardiasis; flagella.  
 XX  
 OS Chlamydomonas sp.  
 PN WO200190307-A2.  
 XX  
 PD 29-NOV-2001.  
 PF 24-MAY-2001; 2001WO-US17103.  
 XX  
 PR 24-MAY-2000; 2000US-206923P.  
 XX  
 PA (UYMA-) UNIV MASSACHUSETTS.  
 XX  
 PI Witman GB, Pazour GJ, Rosenbaum JL, Cole DG;  
 XX  
 DR WPI; 2002-089926/12.  
 DR N-PSDB; ABK27557.  
 XX  
 PT Novel Chlamydomonas or mouse intraflagellar transport (IFT) protein or  
 PT its variant, useful for identifying modulators that are useful for  
 PT treating nematode, insect, protozoa or bacterial infection by  
 PT inhibiting IFT  
 XX  
 PS Claim 7; Figure 12A; 132pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide (I) of Chlamydomonas  
 CC intraflagellar transport (IFT) particle protein, Che-2, or mouse IFT  
 CC particle protein 57. Also described are methods which are useful for:  
 CC diagnosing kidney disease, retinal disorder, thyroid disorder,  
 CC chondrocyte disease, olfactory disease, azoospermia, or primary ciliary

CC dyskinesia; for treating an infection in a mammal (a human) or plant  
 CC caused by nematode, insect, protozoa or bacteria. The nucleic acid  
 CC molecules are useful for diagnosis of disorders associated with aberrant  
 CC expression of nucleic acid molecules and in genetic mapping and  
 CC chromosome identification. The polypeptides are useful in generation of  
 CC antibodies, as reagents in diagnostic assays, for the identification of  
 CC other cellular gene products or compounds that can modulate the activity  
 CC or expression of nucleic acids or polypeptides, and as pharmaceutical  
 CC reagents useful for the treatment of disorders associated with aberrant  
 CC expression or activity of the nucleic acids or polypeptides. IFT can  
 CC serve as a male contraceptive. The IFT proteins can be targeted for  
 CC treating parasitic diseases such as malaria, African sleeping sickness,  
 CC trypanosomiasis, leishmanioses, trichomonosias, and giardiasis. An anti-IFT  
 CC drug, taken orally, would inhibit assembly of flagella in newly divided  
 CC Giardia and cause disassembly of previously formed flagella in non-  
 CC dividing Giardia. An anti-IFT drug would block assembly of the flagellum  
 CC and the flagellar sheath and affect the trypanosome's life cycle. The  
 CC anti-IFT modulators can also be used to treat trichomonos vaginitis in  
 CC humans. IFT inhibitors are also useful for combating phytopathogenic  
 CC nematodes inhibiting insect pest by blocking sensory cilia function  
 CC and/or assembly, leaving an insect unable to smell, unable to taste and  
 CC unable to hear. AAU81494-AAU81529 represent intraflagellar transport  
 CC particle amino acid sequences of the invention.  
 XX  
 SQ Sequence 782 AA;  
 Query Match 3.6%; Score 7; DB 23; Length 782;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 66 DWGNEQL 72  
 DB 769 DWGNEQL 775  
 RESULT 35  
 AAG42436  
 ID AAG42436 standard; Protein; 823 AA.  
 XX  
 AC AAG42436;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 52924.  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.

```

PR 10-JAN-1991; 91US-0639404.
XX (UYTE-) UNIV TENNESSEE.
XX Brian DA, Kapke PA;
XX WPI; 1993-143061/17.
XX DR N-PSDB; AAQ40490.
XX Porcine transmissible gastroenteritis virus DNA coding for
PT nucleocapsid and matrix proteins - used for diagnosis of
PT infection and in prodn. of vaccines
XX
PS Example 1; Columns 17-20; 13pp; English.
XX
CC Clone RG5 was the longest cDNA clone from the 3' end of the
CC porcine TGEV genome to be isolated by the inventors. The largest
CC ORF is predicted to encode a 382 amino acid protein of mol. wt.
CC 43,426. The only TGEV protein described to date that approaches
CC this size is the phosphorylated nucleocapsid protein (46-50 kD by
CC SDS-PAGE. See also AAQ40491.
XX
SQ Sequence 382 AA;
Query Match 3.6%; Score 7; DB 14; Length 382;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 129 RKSRSKS 135
Db 342 RKSRSKS 348
|||||||
|||||||

RESULT 32
ABE77059
ID ABB77059 standard; Protein; 458 AA.
XX
AC ABB77059;
XX
DT 08-OCT-2002 (first entry)
XX
DE Human protein sequence encoded by gene no:17: from cDNA clone HKAEL28.
XX
KW Human; HKAEL28; secreted protein; immunosuppressive; food preservative;
KW antarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; ophthalmological; vulnery; gene therapy; ELISA;
KW radioimmuncassay; enzyme linked immunosorbent assay; autoimmune disease;
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW cerebrovascular disorder; nervous system disorder; ocular disorder;
KW wound healing; food additive.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 403 /label= Unknown
FT FT /label= 455
FT FT /label= Unknown
XX
PN WO200222638-A1.
XX
PD 21-MAR-2002.
XX
PF 17-JAN-2001; 2001WO-US01386.
XX
PR 12-SEP-2000; 2000US-232104P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
PI NI J;

WPI: 2002-258041/30.
XX
XX New nucleic acid molecules encoding 22 human secreted proteins for
XX diagnosing or treating e.g. autoimmune diseases, hyperproliferative
XX disorders, and cardiovascular disorders, and used as food additives or
XX preservatives -
XX
XX Disclosure; Page 49; 526pp; English.
XX
XX The sequence represents a protein sequence of the invention, encoded by
XX cDNA isolated from human clone ID HKAEL28. The invention relates to novel
XX isolated nucleic acid molecules encoding 22 human secreted proteins. The
XX proteins of the invention have immunosuppressive, antiarthritic,
XX antirheumatic, antiproliferative, cytostatic, cardiant, vasotropic,
XX cerebroprotective, nootropic, neuroprotective, antibacterial, virucide,
XX fungicide, ophthalmological, and vulnerary activity. The polynucleotides
XX may have a use in gene therapy. The polynucleotides and polypeptides
XX encoded by them are used to prevent, treat or ameliorate a medical
XX condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
XX chickens or sheep. The polynucleotides and polypeptides are also used in
XX diagnosing a pathological condition or susceptibility to a pathological
XX condition. The antibodies to the proteins can also be used in alleviating
XX symptoms associated with the disorders and in diagnostic immunoassays
XX e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
XX Disorders which are diagnosed or treated include autoimmune diseases,
XX hyperproliferative disorders, cardiovascular disorders, cerebrovascular
XX disorders, angiogenesis, nervous system disorders, infections caused by
XX bacteria, viruses and fungi and ocular disorders. The polypeptides can
XX also be used to aid wound healing and epithelial cell proliferation. The
XX polypeptides can also be used as a food additive or preservative.
XX
SQ Sequence 458 AA;
Query Match 3.6%; Score 7; DB 23; Length 458;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 129 RKSRSKS 135
Db 58 RKSRSKS 64
|||||||
|||||||

RESULT 33
ABE54651
ID ABB54651 standard; Protein; 628 AA.
XX
AC ABB54651;
XX
DT 16-MAY-2002 (first entry)
XX
DE Lactococcus lactis protein pabb.
XX
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
XX WPI; 2002-043418/06.
XX
XX New nucleotide sequence useful in the identification of Lactococcus
XX lactis and related species -
PT
```

CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 334 AA;

Query Match 3.6%; Score 7; DB 21; Length 334;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135  
|||||||  
Db 42 RKSRKS 48

RESULT 29  
AAB51364  
ID AAB51364 standard; Protein; 376 AA.

XX AAB51364;  
XX  
XX 10-APR-2001 (first entry)

XX Feline infectious peritonitis virus nucleocapsid protein SEQ ID NO:2.

XX Feline infectious peritonitis virus; FIP; FIPV; nucleocapsid protein;  
KW vaccine; coronavirus infectious disease.

XX Feline infectious peritonitis virus.

XX JP2000302692-A.

XX 31-OCT-2000.

XX 21-APR-1999; 99JP-0114115.

XX 21-APR-1999; 99JP-0114115.

XX (KYOR-) KYORITSU SHOJI KK.

XX WPI: 2001-141402/15.

XX N-PSDB; AAF26980.

XX A new DNA vaccine against coronavirus infectious disease

XX Example; Page 20; 29pp; Japanese.

XX The present invention describes a DNA vaccine against a coronavirus  
CC infectious disease which comprises a DNA encoding the nucleocapsid  
CC protein (I) in feline infectious peritonitis (FIP) virus and an  
CC expression vector which can express (I) in an applied animal and  
CC physiological saline water. The DNA vaccine can be used for the  
CC prevention and the treatment of a coronavirus infectious disease such  
CC as FIP virus (FIPV). The present sequence represents the FIPV  
CC nucleocapsid protein.

XX Sequence 376 AA;

Query Match 3.6%; Score 7; DB 22; Length 376;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135  
|||||||  
Db 337 RKSRKS 343

RESULT 30  
AAP82001

ID AAP82001 standard; protein; 380 AA.  
XX  
XX AAP82001;  
XX  
XX 14-DEC-1990 (first entry)  
DT  
XX Sequence of phenylalanine dehydrogenase (PADH) encoded by gene isolated  
DE from Bacillus sphaericus SCRC-R79a (FERM BP-1013).  
DE Enzyme.  
XX  
XX Bacillus sphaericus SCRC-R79a (FERM BP-1013).  
OS  
XX EP256514-A.  
PN  
XX 24-FEB-1988.  
PD  
XX 12-AUG-1987; 87EP-0111696.  
PF  
XX 27-NOV-1986; 86JP-0280654.  
PR  
XX 02-AUG-1986; 86JP-0187852.  
PR  
XX (SAGA ) SAGAMI CHEM RES CENTRE.

XX Hibino Y, Asano Y, Okazaki N, Numao N;

XX WPI: 1988-051314/08.

XX N-PSDB; AAN82009.

XX Isolated gene for bacterial phenylalanine dehydrogenase -  
PT with derived plasmids and transformant(s), in phenylalanine  
PT prodn. from phenyl-pyruvic acid

XX Disclosure; 3-3; 50pp; English.

XX The PADH gene (AAN82009) is used to transform microorganisms, esp.  
CC E.coli. Culturing the transformed organisms will produce PADH which is  
CC used to convert phenylpyruvic acid (PPA, or its salt), and NH4 ions (in  
CC presence of a reducing agent) to L-phenylalanine.

XX Sequence 380 AA;

Query Match 3.6%; Score 7; DB 9; Length 380;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 IVKSDDI 32  
|||||||

Db 235 IVKSDDI 241

RESULT 31

AAR39224

ID AAR39224 standard; Protein; 382 AA.

XX AAR39224;

XX 09-AUG-1993 (first entry)

XX Nucleocapsid protein deduced from TGEV clone FG5.

XX Transmissible gastroenteritis virus; pig; vaccine; N protein.

XX Transmissible gastroenteritis Virus.

XX US5202430-A.

XX 13-APR-1993.

XX 23-APR-1987; 87US-0041704.

XX 23-APR-1987; 87US-0041704.

XX 16-JAN-1990; 90US-0465570.



```
Db      275 SKSQSSK 281
|||||||
RESULT 25
AAW73507
ID      AAW73507 standard; Protein; 317 AA.
XX
AC      AAW73507;
XX
DT      01-MAR-1999 (first entry)
XX
DE      Human ATG-1709 protein.
XX
KW      Human; ATG-1709 protein; secreted ligand; 7-Transmembrane receptor;
KW      heart disease; hypertension; cardiovascular disease; kidney disease;
KW      obesity; insulin resistance; diabetes; Central Nervous System disorder;
KW      therapy; sFRP-1.
XX
OS      Homo sapiens.
XX
PN      EP879885-A1.
XX
PD      25-NOV-1998.
XX
PF      16-JAN-1998; 98EP-0300313.
XX
PR      08-AUG-1997; 97US-0907808.
PR      23-MAY-1997; 97US-0047691.
XX
PA      (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI      Hu E, Zhu Y;
XX
WPI: 1998-596877/51.
DR      N-PSDB; AAV08946.
XX
New human secreted protein ATG-1709 polypeptide and polynucleotide -
PT      useful as diagnostic reagents and for diagnosing, prevention and
PT      treatment of Central Nervous System diseases and diabetes
XX
Claim 11; Page 6; 28pp; English.
XX
This sequence represents the human ATG-1709 protein of the invention.
CC      ATG-1709 is related to human secreted ligands for 7-Transmembrane
CC      receptors and similar to murine sFRP-1. ATG-1709 polypeptides and
CC      polynucleotides are useful for diagnosing susceptibility to diseases by
CC      detecting mutations in the ATG-1709 gene using probes containing the
CC      ATG-1709 nucleotide sequence, and can diagnose diseases associated with
CC      ATG-1709 imbalance by determining ATG-1709 polypeptide expression levels.
CC      ATG-1709 polypeptides can be used to screen for agonists and antagonists
CC      which bind the ATG-1709 polypeptide. These can be used in treatment to
CC      activate or inhibit ATG-1709 activity, in addition to direct
CC      administration of antisense sequences to prevent expression, or ATG-1709
CC      polypeptides to treat conditions associated with a lack of ATG-1709.
CC      Gene therapy may also be used to affect endogenous ATG-1709 expression.
CC      ATG-1709 antibodies are useful for inducing an immune response to
CC      immunise and prevent diseases, and for isolating ATG-1709 clones or
CC      purifying the polypeptides by affinity chromatography. ATG-1709
CC      polypeptides can be administered directly or as a vaccine to inoculate
CC      against disease. Diseases diagnosed, prevented or treated include:
CC      heart disease; hypertension; cardiovascular diseases; kidney diseases;
CC      obesity; insulin resistance; diabetes and Central Nervous System (CNS)
CC      diseases. The ATG-1709 polypeptide is also useful for mapping the gene to
CC      a chromosome, allowing gene inheritance to be studied through linkage
CC      analysis.
XX
SQ      Sequence 317 AA;
Query Match 3.6%; Score 7; DB 19; Length 317;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      159 KRKDTKR 165
|||||||
Db      236 KRKDTKR 242
|||||||
RESULT 26
AAW37816
ID      AAW37816 standard; Protein; 317 AA.
XX
AC      AAW37816;
XX
DT      28-AUG-1998 (first entry)
XX
DE      Human secreted apoptosis-related protein hSARP3.
XX
KW      Secreted apoptosis-related protein; SARP; hSARP3; human;
KW      prostate cancer; breast cancer; diagnosis; gene therapy.
XX
OS      Homo sapiens.
XX
FH      Key Location/Qualifiers
FT      Peptide 1..28
FT      /label= Sig_peptide
FT      Protein 29..317
FT      /label= Mat_protein
XX
WO9813493-A2.
XX
PD      02-APR-1998.
XX
PF      24-SEP-1997; 97WO-US17154.
XX
PR      11-OCT-1996; 96US-0028363.
PR      24-SEP-1996; 96US-0026603.
XX
PA      (LXRB-) LXR BIOTECHNOLOGY INC.
XX
PI      Melkonyan H, Umansky S;
XX
WPI: 1998-230704/20.
DR      N-PSDB; AAV19114.
XX
New secreted apoptosis-related proteins - useful for modulating
PT      apoptosis, particularly for treatment of prostatic or breast cancer,
PT      also for diagnosis and monitoring of disease
XX
Claim 1; Page 56-58; 101pp; English.
XX
This polypeptide comprises human secreted apoptosis-related
CC      protein hSARP3 that modulates apoptosis through cell-cell or
CC      cell-extracellular matrix signalling. Its amino acid sequence was
CC      deduced from a clone (see AAV19114) obtained from a human pancreas
CC      cDNA library. Murine mSARP1 (see AAW37814), human hSARP1 (see
CC      AAW37816) and hSARP2 (see AAW37817) proteins are also claimed. hSARP3
CC      is expressed predominantly in pancreas. SARP polypeptides can be
CC      obtained from recombinant host cells. Antibodies specific for SARP
CC      polypeptides can be used in immunoassays for detecting levels of
CC      expression of SARP, particularly for diagnosis or monitoring of
CC      diseases associated with SARP expression. Specifically, they are
CC      used to detect cancer of the prostate or breast (by detecting
CC      hSARP1 and 2, respectively). SARP polypeptides and nucleic acids
CC      can also be used to treat these cancers, or more generally
CC      apoptosis-related disease (e.g. infection with HIV or reperfusion
CC      injury), also (not claimed) to prevent apoptosis in cultured cells,
CC      to improve preservation of organs for transplantation, for in situ
CC      preservation for by-pass operations and to treat dermatological
CC      disorders. SARP polypeptides can also be used to identify agents,
CC      potentially useful therapeutically, that modulate the effects of
CC      SARP on Wnt-frizzled protein interaction.
XX
SQ      Sequence 317 AA;
Query Match 3.6%; Score 7; DB 19; Length 317;
```

Db 260 YFVLENP 266

Qy 107 YFVLENP 113  
1111111  
Db 263 YFVLENP 269

RESULT 24  
AAY41211  
ID AAY41211 standard; Protein; 284 AA.  
XX  
AC AAY41211;  
XX  
DT 31-JAN-2000 (first entry)  
XX  
DE H. influenzae YigT protein.  
XX  
KW Membrane targeting protein; translocation protein; Escherichia coli;  
KW Sec-independent pathway; protein transport; twin-arginine signal peptide;  
KW mtABC operon; MTA protein; YigT protein.  
XX  
OS Haemophilus influenzae.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 110  
FT /note= "unspecified"  
XX  
PN WO9951753-A1.  
XX  
PD 14-OCT-1999.  
XX  
PF 29-MAR-1999; 99WO-CA00272.  
XX  
PR 01-APR-1998; 98US-0053197.  
PR 28-MAY-1998; 98US-0085761.  
XX  
PA (UYAL-) UNIV ALBERTA.  
XX  
PI Weiner JH, Turner RJ;  
XX  
DR WPI; 1999-633740/54.  
XX  
PT New recombinant membrane targeting and translocation proteins from  
PT Escherichia coli, used to produce soluble polypeptides normally  
XX produced in insoluble form -  
PS Disclosure; Fig 6; 111pp; English.  
XX  
CC The invention relates to recombinant membrane targeting and translocation  
CC proteins from Escherichia coli. The recombinant polypeptides are involved  
CC in the Sec-independent pathway for transporting proteins with a twin-  
CC arginine signal peptide to the periplasm or extracellular medium and to  
CC the cell membrane. They transport fully folded and co-factor containing  
CC proteins. The polypeptides are used to transport such proteins, normally  
CC specifically to produce soluble forms of polypeptides that are, normally  
CC produced in insoluble form. They may also be used to raise specific  
CC antibodies. Nucleic acid sequences that encode the polypeptides are used  
CC for production of recombinant proteins and their fragments are used as  
CC probes to detect or isolate related genomic or cDNA sequences (these have  
CC been found in many other bacteria, yeast, plants, nematodes and humans).  
CC Producing normally insoluble proteins in soluble form eliminates the need  
CC for resolubilization with strong denaturants and facilitates recovery of  
CC functional proteins (which have been properly folded by cytoplasmic  
CC enzymes before translocation). The present sequence represents the  
CC H. influenzae YigT protein which has homology to the E. coli MTA  
CC protein.  
XX  
SQ Sequence 284 AA;

Query Match 3.6%; Score 7; DB 20; Length 284;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 133 SKSQSSK 139

Db 260 YFVLENP 266

RESULT 23  
AAM41951  
ID AAM41951 standard; Protein; 283 AA.  
XX  
AC AAM41951;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 6882.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
XX  
DR N-PSDB; AAI61107.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 2; SEQ ID NO 6882; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 283 AA;

Query Match 3.6%; Score 7; DB 22; Length 283;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

RESULT 21
AAU30761
ID AAU30761 standard; Protein; 266 AA.
XX
AC AAU30761;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #1252.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 341; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 266 AA;
Query Match 3.6%; Score 7; DB 22; Length 266;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 RPKRRK 130
| | | | |
Db 157 RPKRRK 163

RESULT 22
AAB42139
ID AAB42139 standard; Protein; 280 AA.
XX
AC AAB42139;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1903 polypeptide sequence SEQ ID NO:3806.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC76348.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 2956; 5507pp; English.
XX
AAAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 280 AA;
Query Match 3.6%; Score 7; DB 21; Length 280;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 YFVLENP 113
| | | | |

```



PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.6%; Score 7; DB 21; Length 256;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTPEM 9  
| | | | |  
Db 117 KEVTPEM 123

RESULT 19  
AAB36598  
ID AAB36598 standard; Protein; 254 AA.  
XX  
AC AAB36598;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human FLEXHT-20 protein sequence SEQ ID NO:20.  
XX  
KW Human; FLEXHT: full-length molecules expressed in human tissue;  
KW diagnosis; gene expression; genetic linkage; genetic variability;  
KW antianemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;  
KW cytoskeletal; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;  
KW anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;  
KW antipsoriatic; antirheumatic; antiulcer; gene therapy; anaemia; gout;  
KW epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;  
KW cancer; immunological disorder; asthma; bronchitis; cirrhosis;  
KW Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;  
KW osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;  
KW ulcerative colitis.  
XX  
OS Homo sapiens.  
XX  
PN WO200070047-A2.  
XX  
PD 23-NOV-2000.  
XX  
PF 12-MAY-2000; 2000WO-US13299.  
XX  
PR 14-MAY-1999; 99US-0311894.  
PR 14-MAY-1999; 99US-0311937.  
PR 14-MAY-1999; 99US-0311940.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Yue H, Tang YT, Lal P, Reddy R, Batra S, Baughn MR, Yang J;  
PI Azimzai Y, Lu DAM, Au-Young J, Shih LL;  
XX  
WPI: 2001-016234/02.  
DR N-PSDB; AAC88089.  
XX  
PT Human FLEXHT protein and DNA sequences, useful for treating  
PT immunological disorders, developmental disorders, and cancers -  
PS Claim 1; Page 113; 168pp; English.  
XX  
CC AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules  
CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The  
CC present invention describes an isolated polypeptide (A) comprising an  
CC amino acid sequence selected from one of 55 amino acid sequences 42-876  
CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %  
CC identical sequence, and a biologically active or immunogenic fragment of  
CC the sequence. The FLEXHT proteins can have antianemic, anticonvulsant,  
CC antiarteriosclerotic, immunomodulatory, cytoskeletal, antiasthmatic,  
CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antithyroid,  
CC neuroprotective, antiarthritic, osteopathic, antipsoriatic, antiulcer  
CC and antirheumatic activities, and can be used in gene therapy. The  
CC polynucleotide sequences can be used to express the protein sequences.  
CC Pharmaceutical compositions comprising FLEXHT can be used to treat  
CC diseases or conditions associated with altered expression of functional  
CC FLEXHT. The proteins and polynucleotides can be used to diagnose and  
CC treat disorders including anaemia, epilepsy, arteriosclerosis,  
CC atherosclerosis, developmental disorders, cancers, and immunological  
CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,  
CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,  
CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and  
CC ulcerative colitis.  
XX  
SQ Sequence 254 AA;  
Query Match 3.6%; Score 7; DB 22; Length 254;  
Best Local Similarity 100.0%; Pred. No. 45;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 107 YFVLENP 113  
| | | | | | | |  
Db 234 YFVLENP 240  
RESULT 20  
AAG09869  
ID AAG09869 standard; Protein; 256 AA.  
XX  
AC AAG09869;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7968.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140595.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142520.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.6%; Score 7; DB 21; Length 248;  
Best Local Similarity 100.0%; Pred No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTPM 9  
|||||  
Db 109 KEVTPM 115

```
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.6%; Score 7; DB 21; Length 240;
Best Local Similarity 100.0%; Pred. No. 43;
```

```
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 KEVTPEM 9
Db 101 KEVTPEM 107

RESULT 18
AAG09870
ID AAG09870 standard; Protein; 248 AA.
XX
AC AAG09870;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7969.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
```

CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX  
SQ Sequence 235 AA;

Query Match 3.6%; Score 7; DB 22; Length 235;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 YFVLENP 113  
|||||||  
Db 215 YFVLENP 221

## RESULT 17

AAG09871  
ID AAG09871 standard; Protein; 240 AA.

XX

AC AAG09871;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 7970.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.  
XX EP1033405-A2.

PN

XX PD 06-SEP-2000.

XX

XX PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.

CC useful for treating chronic neurodegenerative diseases such as  
 CC Alzheimer's disease and Parkinsonian's disease, autoimmune diseases,  
 CC diabetes mellitus, congenital muscular dystrophy with mitochondrial  
 CC structural abnormalities, fatal infantile myopathy, mitochondrial  
 CC encephalopathy, lactic acidosis, stroke, mitochondrial diabetes,  
 CC deafness, neuropathy, progressive external ophthalmoplegia, optic  
 CC atrophy, Leigh's syndrome, dystonia, stroke, schizophrenia, progressive  
 CC joint disorders such as osteoarthritis, and hyperproliferative  
 CC disorders such as cancer, tumour and psoriasis.  
 CC

SQ Sequence 157 AA;

Query Match 3.6%; Score 7; DB 22; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 53 FSEILK 59  
 |||||  
 Db 53 FSEILK 59

#### RESULT 15

AAW93976  
 ID AAW93976 standard; protein; 159 AA.

XX AAW93976;

AC AAW93976;

DT 30-JUN-1999 (first entry)

DE Human IRS-1 and IRS-2 binding inhibitor peptide 4.

XX Insulin receptor substrate; IRS-1; IRS-2; inhibitor; disease; obesity;  
 KW insulin resistance; 14-3-3 protein; inhibitor; screening; diabetes;  
 KW diabetic retinopathy; diabetic neuropathy; glucose tolerance;  
 KW diabetic nephropathy; hyperinsulinaemia; hyperlipemia; arteriosclerosis;  
 KW hypertension; ischaemic heart disease; ischaemic brain disease;  
 KW peripheral embolism.

OS Homo sapiens.

XX W09916462-A1.

PN 08-APR-1999.

XX 25-SEP-1998; 98WO-JP04293.

XX 29-SEP-1997; 97JP-0263719.

XX (DAUC ) DAIICHI PHARM CO LTD.

XX Asano T, Kanda A, Kubo H, Yazaki Y;

XX WPI; 1999-254929/21.

XX Treatment of insulin resistance using insulin receptor substrate  
 PT binding inhibitor

XX Disclosure; Page 24-25; 30pp; Japanese.

CC This invention describes a method for the treatment of diseases involving  
 CC insulin resistance using a substance which inhibits the binding of  
 CC insulin receptor substrate to 14-3-3 protein, identified by screening  
 CC potential inhibitors for their ability to inhibit this binding. The  
 CC composition described in the invention for the treatment of diseases  
 CC involving insulin resistance contains as active component, an inhibitor  
 CC of the binding of all or part of insulin receptor substrate 1 or 2  
 CC (IRS-1, IRS-2) to all or part of 14-3-3 protein. Disorders involving  
 CC insulin resistance include diabetes, diabetic retinopathy, diabetic  
 CC neuropathy, impaired glucose tolerance, diabetic nephropathy,  
 CC hyperinsulinaemia, hyperlipemia, arteriosclerosis, hypertension,  
 CC obesity, ischaemic heart disease, ischaemic brain disease, and peripheral  
 CC embolism.  
 CC

SQ Sequence 159 AA;

Query Match 3.6%; Score 7; DB 20; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 132 RSKSQSS 138  
 |||||

Db 108 RSKSQSS 114

#### RESULT 16

AAW40165

ID AAW40165 standard; Protein; 235 AA.

XX AAW40165;

AC AAW40165;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3310.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX Homo sapiens.

OS W0200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI59321.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -

XX Example 5; SEQ ID NO 3310; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAW38642-AAW42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.

CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX SQ Sequence 129 AA;

Query Match 3.6%; Score 7; DB 22; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105  
 Db 75 EYCNFGC 81  
 |||||

RESULT 13

ABP40049

ID ABP40049 standard; Protein; 132 AA.

XX AC ABP40049;

XX XX

24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4894.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

PR 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

DR N-PSDB; ABN92594.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 PS Disclosure; SEQ ID 4894; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.

XX SQ Sequence 132 AA;

Query Match 3.6%; Score 7; DB 23; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105  
 Db 78 EYCNFGC 84  
 |||||

RESULT 14

AAB47229

ID AAB47229 standard; Protein; 157 AA.

XX AC AAB47229;

XX 18-JUL-2001 (first entry)

XX Yeast Mfn.

XX Mitofusin; homolog; GTPase; mediation; mitochondrial fusion; stroke;  
 KW post-meiotic fusion; mitochondria; spermatid; anti-insect; deafness;  
 KW antifungal; fertility; mortality; arthropod pest; Leigh's syndrome;  
 KW mitochondrial disease; muscle myopathy; neurodegenerative disorder;  
 KW aging; Alzheimer's disease; Parkinsonian's disease; cancer; psoriasis;  
 KW autoimmune disease; diabetes mellitus; congenital muscular dystrophy;  
 KW fatal infantile myopathy; mitochondrial encephalopathy; optic atrophy;  
 KW lactic acidosis; mitochondrial diabetes; neuropathy; osteoarthritis;  
 KW progressive external ophthalmoplegia; dystonia; schizophrenia.

XX Saccharomyces cerevisiae.

XX WO200125274-A1.

XX 12-APR-2001.

XX 06-OCT-2000; 2000WO-US27871.

XX 06-OCT-1999; 99US-0413285.

XX (STRD ) UNIV LELAND STANFORD JUNIOR.

XX Fuller MT, Hales KG, Santel AH;

XX WPI; 2001-300165/31.

XX New nucleic acid molecules encoding mitofusin polypeptides useful for  
 PT identifying anti-insect, antifungal and therapeutic agents -

XX Disclosure; Page 83-84; 92pp; English.

XX The sequences given in AAB47221-30 show mitofusin proteins and their  
 CC homologs. Mitofusins are large predicted GTPases with a predicted  
 CC transmembrane domain, coiled-coil region and a C-terminal region showing  
 CC a high isoelectric point and a predicted coiled-coil region. They are  
 CC the first known protein mediators of mitochondrial fusion, e.g.  
 CC mediating developmentally regulated post-meiotic fusion of mitochondria  
 CC in Drosophila spermatids. Mitofusin polypeptides are useful for  
 CC enhancing membrane fusion in a population of membrane-bounded entities,  
 CC e.g., mitochondria. Mitofusin proteins are useful for screening  
 CC anti-insect or antifungal agent. Modulators of mitochondrial fusion  
 CC are useful for reducing fertility and increasing the mortality rate  
 CC of arthropod pest, e.g., an insect (claimed). Mitofusin DNA is  
 CC useful for generating genetically modified non-human animals or  
 CC site-specific gene modifications in cell lines, for the study of  
 CC mitofusin function or regulation, and to create animal models of  
 CC diseases, including mitochondrial diseases, muscle myopathies,  
 CC neurodegenerative disorders, and aging. Mitofusins, and the DNA  
 CC encoding them, can be used in assays to identify therapeutic agents

XX WO9934015-A2.  
 PN 08-JUL-1999.  
 PD 22-DEC-1998; 98WO-GB03887.  
 PF 27-FEB-1998; 98GB-0004050.  
 PR 24-DEC-1997; 97GB-0027347.  
 XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.  
 PA Ganesan TS;  
 XX WPI: 1999-405519/34.  
 PI N-PSDB; AAX84493.  
 DR Methods of diagnosis, prognosis and treatment of cancer  
 XX Disclosure; Fig 6; 167pp; English.  
 XX This sequence is the p90 ribosomal S6 kinase-3 (Rsk-3).  
 CC The invention relates of diagnosis, prognosis and treatment of cancer  
 CC related to the Rsk-3 gene. The methods are used for diagnosis, prognosis  
 CC and treatment of cancer, especially ovarian or breast cancer or lymphoma.  
 CC Nucleic acids that selectively hybridise to the Rsk-3 gene or cDNA, or a  
 CC mutant Rsk-3 allele, or a molecule that selectively binds to Rsk-3  
 CC polypeptide can be used to manufacture reagents for diagnosis of cancer.  
 CC Rsk-3 can be used to treat or ameliorate cancer. Rsk-3 inhibitors can  
 CC also be used to treat cancer. Vaccines comprising mutant Rsk-3 or nucleic  
 CC acids encoding mutant Rsk-3, where the mutant is found in cancer cells,  
 CC are useful for treatment of cancer.  
 XX SQ Sequence 733 AA;  
 SQ  
 Query Match 4.2%; Score 8; DB 20; Length 733;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 128 RRSRSKS 135  
 DB 19 RRSRSKS 26  
 RESULT 11  
 AAW93973  
 ID AAW93973 standard; protein; 15 AA.  
 AC AAW93973;  
 XX 30-JUN-1999 (first entry)  
 DT Human IRS-1 and IRS-2 binding inhibitor peptide 1.  
 DE Insulin receptor substrate; IRS-1; IRS-2; inhibitor; disease; obesity;  
 XX insulin resistance; 14-3-3 protein; inhibitor; screening; diabetes;  
 KW diabetic retinopathy; diabetic neuropathy; glucose tolerance;  
 KW diabetic nephropathy; hyperinsulinaemia; hyperlipemia; arteriosclerosis;  
 KW hypertension; ischaemic heart disease; ischaemic brain disease;  
 KW peripheral embolism.  
 XX Homo sapiens.  
 OS WO9916462-A1.  
 XX 08-APR-1999.  
 PD 25-SEP-1998; 98WO-JP04293.  
 XX 29-SEP-1997; 97JP-0263719.  
 PR (DAUC ) DAIICHI PHARM CO LTD.  
 PA

PI Asano T, Kanda A, Kubo H, Yazaki Y;  
 XX WPI: 1999-254929/21.  
 DR Treatment of insulin resistance using insulin receptor substrate  
 XX binding inhibitor  
 PT Disclosure; Page 23; 30pp; Japanese.  
 XX  
 CC This invention describes a method for the treatment of diseases involving  
 CC insulin resistance using a substance which inhibits the binding of  
 CC insulin receptor substrate to 14-3-3 protein, identified by screening  
 CC potential inhibitors for their ability to inhibit this binding. The  
 CC composition described in the invention for the treatment of diseases  
 CC involving insulin resistance contains as active component, an inhibitor  
 CC of the binding of all or part of insulin receptor substrate 1 or 2  
 CC (IRS-1, IRS-2) to all or part of 14-3-3 protein. Disorders involving  
 CC insulin resistance include diabetes, diabetic retinopathy, diabetic  
 CC neuropathy, impaired glucose tolerance, diabetic nephropathy,  
 CC hyperinsulinaemia, hyperlipemia, arteriosclerosis, hypertension,  
 CC obesity, ischaemic heart disease, ischaemic brain disease and peripheral  
 CC embolism.  
 XX SQ Sequence 15 AA;  
 SQ  
 Query Match 3.6%; Score 7; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 132 RSKSQSS 138  
 DB 6 RSKSQSS 12  
 RESULT 12  
 AAG82114  
 ID AAG82114 standard; Protein; 129 AA.  
 XX AAG82114;  
 AC AAG82114;  
 XX 03-SEP-2001 (first entry)  
 DT S. epidermidis open reading frame protein sequence SEQ ID NO:1322.  
 XX Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 DE vaccination; endocarditis.  
 KW Staphylococcus epidermidis.  
 KW Staphylococcus epidermidis.  
 XX WO200134809-A2.  
 XX 17-MAY-2001.  
 PD 09-NOV-2000; 2000WO-US30782.  
 PF 09-NOV-1999; 99US-0164258.  
 XX (GLAXO ) GLAXO GROUP LTD.  
 PA Kimmerly WJ;  
 XX WPI: 2001-316495/33.  
 DR N-PSDB; AAH52964.  
 DR Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 XX Claim 18; Page 376-377; 2188pp; English.  
 XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the



PF 29-OCT-2001; 2001WO-GB04789.  
 XX  
 PR 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 PI  
 XX WPI; 2002-352536/38.  
 DR N-PSDB; ABN69287.  
 DR  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX  
 XX Claim 1; Page 3811; 4525pp; English.  
 PS  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and  
 CC antibodies that bind (1) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
 CC biological sample. (1) is used to determine whether a compound binds to  
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (1) may be used to recombinantly produce (1) and may be  
 CC used in gene therapy. Antibodies to (1) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX  
 SQ Sequence 303 AA;  
 Query Match 4.2%; Score 8; DB 23; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 4.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 TVFGQRES 54  
 Db | | | | | | | |  
 27 TVFGQRES 34  
 RESULT 9  
 AAY30937  
 ID AAY30937 standard; Protein; 733 AA.  
 AC  
 XX AAY30937;  
 XX  
 XX 18-OCT-1999 (first entry)  
 DT  
 XX Human rsk3 protein.  
 DE  
 XX CD30; antigen; rsk3; ribosomal S6 kinase; pp90rsk3; cell activation;  
 KW antitumor; antiviral; antibacterial; antiparasitic; anti-allergic;  
 KW anti-inflammatory; antiproliferative; suppressor; cell death; human;  
 KW activation-induced proliferation; viral; tumour cell; chemotherapy;  
 KW cytokine; radiation; treatment; prevention immune dysregulation; blood;  
 KW allergy; inflammatory disease; transplant rejection; germ cell; mucin;  
 KW Hodgkin's lymphoma; anaplastic giant cell leukemia; embryonal carcinoma;  
 KW transplant; bone marrow; cystic fibrosis.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO940187-A1.  
 PN  
 XX

PD 12-AUG-1999.  
 XX  
 XX 05-FEB-1999; 99WO-EP00759.  
 XX  
 PR 22-DEC-1998; 98DE-1059056.  
 PR 06-FEB-1998; 98DE-1038967.  
 XX  
 XX (ABKE/) ABKEN H.  
 PA  
 XX Abken H;  
 PI  
 XX WPI; 1999-494292/41.  
 DR N-PSDB; AAZ09173.  
 DR  
 XX  
 PT New nucleic acids that inhibit expression of the CD30 antigen or  
 PT ribosomal kinase rsk3, for modulating activity of, or killing, cells  
 PT expressing these proteins, e.g. tumor cells  
 XX  
 XX Claim 10; Page 45-48; 60pp; German.  
 PS  
 XX This invention describes novel nucleic acids (I), their derivatives and  
 CC sequence-specific binding substances that inhibit expression of the CD30  
 CC antigen or the ribosomal S6 kinase pp90rsk3 (rsk3). The products of the  
 CC invention have antitumor, antiviral, antibacterial, antiparasitic,  
 CC anti-allergic, anti-inflammatory and antiproliferative activity. (I)  
 CC suppress expression of CD30 and/or rsk3 by specific binding to related  
 CC nucleic acid, so prevent activation of CD30/rsk3-positive cells, restore  
 CC deregulated production of cellular products (particularly cytokines or  
 CC other effectors), suppress activation-induced proliferation and viral  
 CC production, induce cell death and may render tumour cells sensitive to  
 CC killing by chemotherapeutics, cytokines or radiation. (I) are used to  
 CC (1) modulate activity of CD30 and/or rsk3-positive cells, and (2) for  
 CC functional suppression or killing of such cells. Typical applications  
 CC are: treatment/prevention of immune dysregulation (e.g. in cases of  
 CC allergy, inflammatory disease, transplant rejection); reduction of viral  
 CC load (e.g. in infections by human immune deficiency virus, Epstein-Barr  
 CC virus, hepatitis); elimination of tumours (many of which express CD30,  
 CC e.g. Hodgkin's lymphoma, anaplastic giant cell leukemia, germ cell  
 CC tumours and embryonal carcinomas), and treatment of viral, bacterial and  
 CC parasitic infections. They may also be used to suppress/kill such cells  
 CC in transplants, bone marrow samples, also in blood (for treatment of  
 CC allergies), and further (I) directed against rsk3 inhibit mucin  
 CC overexpression induced by Pseudomonas aeruginosa, e.g. in cases of cystic  
 CC fibrosis. (I) selectively inhibit, or kill, CD30 and/or rsk3-positive  
 CC cells without significant harm to other cells or tissues. This  
 CC sequence represents the human rsk3 protein which is used in the method of  
 CC the invention.  
 XX  
 XX Sequence 733 AA;  
 SQ  
 Query Match 4.2%; Score 8; DB 20; Length 733;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 128 RRKSRSKS 135  
 Db | | | | | | | |  
 19 RRKSRSKS 26  
 RESULT 10  
 AAY22202  
 ID AAY22202 standard; Protein; 733 AA.  
 XX  
 XX AAY22202;  
 AC  
 XX 13-SEP-1999 (first entry)  
 DT  
 XX p90 ribosomal S6 kinase-3 protein sequence.  
 DE  
 XX Rsk-3; p90 ribosomal S6 kinase-3; diagnosis; therapy; ovarian cancer;  
 KW breast cancer; lymphoma; cancer; vaccine.  
 KW  
 XX Homo sapiens.  
 OS

ID ABB50067 standard; Protein: 91 AA.  
 XX ABB50067;  
 AC  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE Listeria monocytogenes protein #2771.  
 XX  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 XX  
 OS Listeria monocytogenes.  
 XX  
 PN WO200177335-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-FR01118.  
 XX  
 PR 11-APR-2000; 2000FR-0004629.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX  
 WPI: 2002-010914/01.  
 XX  
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides -  
 XX  
 PS Claim 6; SEQ ID No 2772; 192pp; French.  
 XX  
 CC The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 91 AA;  
 Query Match 5.7%; Score 11; DB 23; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 0.0008;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 99 EYCNFGCAFYV 109  
 Db 76 EYCNFGCAFYV 86  
 RESULT 7  
 ABB53482  
 ID ABB53482 standard; Protein: 251 AA.

XX ABB53482;  
 AC  
 XX  
 DT 16-MAY-2002 (first entry)  
 XX  
 DE Lactococcus lactis protein ybiC.  
 XX  
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
 XX  
 OS Lactococcus lactis IL1403.  
 XX  
 PN FR2807446-A1.  
 XX  
 PD 12-OCT-2001.  
 XX  
 PF 11-APR-2000; 2000FR-0004630.  
 XX  
 PR 11-APR-2000; 2000FR-0004630.  
 XX  
 PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
 XX  
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
 XX  
 WPI: 2002-043418/06.  
 XX  
 PT New nucleotide sequence useful in the identification or Lactococcus  
 PT lactis and related species -  
 XX  
 PS Claim 6; SEQ ID No 184; 2504pp; French.  
 XX  
 CC The present invention is related to a Lactococcus lactis nucleotide  
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC WO200177334 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 251 AA;  
 Query Match 4.2%; Score 8; DB 23; Length 251;  
 Best Local Similarity 100.0%; Pred. No. 3.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 10 LNYNKYPG 17  
 Db 10 LNYNKYPG 17  
 RESULT 8  
 ABP28656  
 ID ABP28656 standard; Protein: 303 AA.  
 XX  
 AC ABP28656;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 6488.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX

```
RESULT 4
AAU37974
ID AAU37974 standard; Protein; 176 AA.
XX AC AAU37974;
XX DT 14-FEB-2002 (first entry)
XX DE Streptococcus pneumoniae cellular proliferation protein #403.
XX KW Antisense; prokaryotic cellular proliferation protein;
XX KW antibiotic; antibacterial; drug design.
XX OS Streptococcus pneumoniae.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253623P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX DR WPI; 2001-611495/70.
XX DR N-PSDB; AAS55833.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS Example 3; Seq ID No 13567; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence represents an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 176 AA;
Query Match 6.2%; Score 12; DB 22; Length 176;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 NEQLRLRGFYKD 80
DB 69 NEQLRLRGFYKD 80
|||||
RESULT 5
ABP25487
ID ABP25487 standard; Protein; 214 AA.
XX AC ABP25487;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 150.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus pyogenes.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB04789.
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN66118.
XX PT New Streptococcus protein for the treatment or prevention of infection
XX PT or disease caused by Streptococcus bacteria, such as meningitis, and
XX PT for detecting a compound that binds to the protein -
XX PS Claim 1; Page 3168; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins.
XX SQ Sequence 214 AA;
Query Match 6.2%; Score 12; DB 23; Length 214;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 63 IVGDWGNQEQLRL 74
DB 71 IVGDWGNQEQLRL 82
|||||
RESULT 6
ABB50067
```

DE Streptococcus polypeptide SEQ ID NO 148.

XX Streptococcus; GAS; GBs; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

KW antinflammatory; infection; vaccine; meningitis; gene therapy.

XX

XX Streptococcus agalactiae.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

XX

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tettelin H;

XX WPI; 2002-352536/38.

DR N-PSDB; ABN66117.

XX

PT New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

XX

PS Claim 1; Page 3168; 4525pp; English.

XX

CC The invention relates to a protein (ABP25413-ABP30895) from group B

CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

CC the specification. The proteins have antibacterial and antinflammatory

CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

XX

SQ Sequence 192 AA;

Query Match 96.4%; Score 185; DB 23; Length 192;

Best Local Similarity 100.0%; Pred. No. 3.3e-192;

Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRKEVTPMLNKNYPGQGFHFENIVKSDDELFOLVINEKSAFDVTVFGQRFSEILLKY 60

Db 1 MRKEVTPMLNKNYPGQGFHFENIVKSDDELFOLVINEKSAFDVTVFGQRFSEILLKY 60

Qy 61 DFIVGDWNEQLRLRGFYKDASTIRKNSRISRLDEYIKECNFGCAYFVLEPNPNRDIKF 120

Db 61 DFIVGDWNEQLRLRGFYKDASTIRKNSRISRLDEYIKECNFGCAYFVLEPNPNRDIKF 120

Qy 121 DDERPHKRRKSKSQSKSQSOTNRNRSQSNANAHFTSKRKDKTKRQRERHKEQDKWMT 180

Db 121 DDERPHKRRKSKSQSKSQSOTNRNRSQSNANAHFTSKRKDKTKRQRERHKEQDKWMT 180

Qy 181 SAKQH 185

Db 181 SAKQH 185

RESULT 3

AAU37821

ID AAU37821 standard; Protein; 176 AA.

XX

AC AAU37821;

XX

DT 14-FEB-2002 (first entry)

XX

DE Streptococcus pneumoniae cellular proliferation protein #250.

XX

KW Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX

OS Streptococcus pneumoniae.

XX WO200170955-A2.

XX

PD 27-SEP-2001.

XX

PF 21-MAR-2001; 2001WO-US09180.

XX

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX

DR WPI; 2001-611495/70.

XX N-PSDB; AAS55680.

XX

PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX

Example 3; Seq ID No 13414; 511pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 176 AA;

Query Match 6.2%; Score 12; DB 22; Length 176;

Best Local Similarity 100.0%; Pred. No. 0.00012;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 NEQLRLRGFYKD 80

Db 69 NEQLRLRGFYKD 80

960 5 2.6 44 22 ABB30911 Peptide #3562 enco  
961 5 2.6 44 22 ABB35277 Peptide #2783 enco  
962 5 2.6 44 22 ABB35386 Peptide #2892 enco  
963 5 2.6 44 22 ABB36094 Peptide #3600 enco  
964 5 2.6 44 22 ABB20720 Protein #2719 enco  
965 5 2.6 44 22 ABB20826 Protein #2823 enco  
966 5 2.6 44 22 ABB21485 Protein #3484 enco  
967 5 2.6 44 22 AAM95781 Human reproductive  
968 5 2.6 44 22 AAM56107 Human brain expres  
969 5 2.6 44 22 AAM56215 Human brain expres  
970 5 2.6 44 22 AAM56876 Human brain expres  
971 5 2.6 44 22 AAM68478 Human bone marrow  
972 5 2.6 44 22 AAM68591 Human bone marrow  
973 5 2.6 44 22 AAM69265 Human immune/haema  
974 5 2.6 44 22 AAM8275 Human polypeptide  
975 5 2.6 44 22 AAC02529 Human polypeptide  
976 5 2.6 44 22 AAC02779 Peptide #2720 enco  
977 5 2.6 44 22 AAM16286 Peptide #2832 enco  
978 5 2.6 44 22 AAM16398 Peptide #3529 enco  
979 5 2.6 44 22 AAM17095 Peptide #2812 enco  
980 5 2.6 44 22 AAM28775 Peptide #2932 enco  
981 5 2.6 44 22 AAM28895 Peptide #3622 enco  
982 5 2.6 44 22 AAM29585 Peptide #2701 enco  
983 5 2.6 44 22 AAM04019 Peptide #2816 enco  
984 5 2.6 44 22 AAM04134 Peptide #3475 enco  
985 5 2.6 44 22 AAM04793 Human peptide enco  
986 5 2.6 44 23 ABG38060 Human peptide enco  
987 5 2.6 44 23 ABG38168 Human peptide enco  
988 5 2.6 44 23 ABG38878 Human peptide enco  
989 5 2.6 45 20 AAY26052 Human Zalphal epit  
990 5 2.6 45 21 AAB16574 Bacteriophage 44AH  
991 5 2.6 45 21 ABG01448 Arabidopsis thalia  
992 5 2.6 45 21 AAG34478 Arabidopsis thalia  
993 5 2.6 45 21 AAY65755 Breast cancer susc  
994 5 2.6 45 22 ABB37435 Peptide #4941 enco  
995 5 2.6 45 22 ABB22734 Protein #4733 enco  
996 5 2.6 45 22 AAU20159 Human novel endocr  
997 5 2.6 45 22 AAM84395 Human immune/haema  
998 5 2.6 45 22 AAB66328 D radiolurans glut  
999 5 2.6 45 23 ABB81765 Human alpha helica  
1000 5 2.6 46 11 AAR05208 Ubiquitin C-termin

ALIGNMENTS

RESULT 1  
AA996805  
ID AAY96805 standard; Protein; 192 AA.  
XX AC AAY96805;  
XX 26-SRP-2000 (first entry)  
XX Streptococcus agalactiae pho3-1 protein.  
DE DE  
XX KW Group B Streptococcus; pho3-1; virulence; vaccine; uropathic;  
KW gene therapy; veterinary; mastitis; immunostimulant; antibacterial.  
XX OS Streptococcus agalactiae.  
XX PN WQ200037646-A2.  
XX PD 29-JUN-2000.  
XX PF 22-DEC-1999; 99WO-GB04377.  
XX PR 22-DEC-1998; 98GB-0028345.  
PR 22-DEC-1998; 98GB-0028349.  
PR 22-DEC-1998; 98GB-0028350.  
PR 22-DEC-1998; 98GB-0028352.  
PR 22-DEC-1998; 98GB-0028353.  
PR 22-DEC-1998; 98GB-0028354.  
PR 22-DEC-1998; 98GB-0028355.  
PR 22-DEC-1998; 98GB-0028356.  
PR 22-DEC-1998; 98GB-0028357.  
PR 22-DEC-1998; 98GB-0028359.  
PR 04-JAN-1999; 98GB-0000082.  
PR 04-JAN-1999; 98GB-0000083.  
PR 04-JAN-1999; 98GB-0000084.  
PR 04-JAN-1999; 98GB-0000085.  
PR 04-JAN-1999; 98GB-0000086.  
PR 28-JAN-1999; 98GB-0001916.  
PR 28-JAN-1999; 98GB-0001922.  
XX (MICR-) MICROSCIENCE LTD.  
PA Hughes MJG, Santangelo JD, Lane JD, Everest P, Feldman R;  
XX Moore JC, Wilson RK, Dobson RJ, Dougan G;  
XX WPI: 2000-442674/38.  
XX N-PSDB; AAA51368.  
XX New peptides useful for treatment and prevention of conditions  
PT associated with Streptococcal infection are obtained from group B  
PT Streptococcus  
XX PS  
XX Claim 2; Page 47-48; 63pp; English.  
XX The S. agalactiae pho3-1 protein has homology with proteins of unknown  
CC function in S. pyogenes, S. pneumoniae, Bacillus subtilis (yutD) and  
CC Enterococcus faecalis. The B. subtilis yutD gene is located in a  
CC chromosomal region containing genes involved in cell wall synthesis.  
CC Peptides derived from group B Streptococcus (Streptococcus agalactiae)  
CC proteins encoded by genes pho1-13, pho3-21, pho3-15, pho3-18, pho3-22,  
CC pho3-3, pho3-17, pho3-2, pho1-5, pho3-1, pho3-23, pho3-50, pho1-14,  
CC pho2-10, pho3-14, pho3-24 and pho3-29. The peptides are useful for  
CC screening potential drugs, or for the detection of virulence, and for the  
CC manufacture of a medicament for use in the treatment or prevention of the  
CC infections such as focal infection (including osteomyelitis, septic  
CC arthritis, abscesses and endophthalmitis) and urinary tract infections  
CC caused by group B Streptococcus. The peptides and vaccines comprising the  
CC peptides are useful in treatment of chronic mastitis, especially in cows  
CC (i.e. for veterinary purposes).  
XX SQ  
SQ Sequence 192 AA;  
Query Match 100.0%; Score 192; DB 21; Length 192;  
Best Local Similarity 100.0%; Pred. No. 8.4e-200;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRKEVTPPEMLNKNYKPGQFIHFENIVKSDDIIEFQLVINEKSAFDVTVFGQRFSEILKY 60  
DB 1 MRKEVTPPEMLNKNYKPGQFIHFENIVKSDDIIEFQLVINEKSAFDVTVFGQRFSEILKY 60  
QY 61 DFIVGDWGNQRLRGFYKDASTIRKNSRISRLIEDYIKEYCNFCGAYFVLENNPRDIKF 120  
DB 61 DFIVGDWGNQRLRGFYKDASTIRKNSRISRLIEDYIKEYCNFCGAYFVLENNPRDIKF 120  
QY 121 DDERPHKRRKRSKSSQSSKTSQTRNNRSQSNAHAFTSKRKDKTKRQQRHIKEEQDKEMT 180  
DB 121 DDERPHKRRKRSKSSQSSKTSQTRNNRSQSNAHAFTSKRKDKTKRQQRHIKEEQDKEMT 180  
QY 181 SAKQHLLFVRKN 192  
DB 181 SAKQHLLFVRKN 192  
RESULT 2  
ABP25486  
ID ABP25486 standard; Protein; 192 AA.  
XX AC ABP25486;  
XX DT 02-JUL-2002 (first entry)  
XX

814	5	2.6	24	23	ABB76466	Mouse IgA inducing	887	5	2.6	36	22	AAE06927	Peptide comprising
815	5	2.6	24	23	ABG42626	Human peptide enco	888	5	2.6	37	20	AAV12015	Human 5' EST secre
816	5	2.6	25	20	AAV27521	E. coli beta'-subu	889	5	2.6	37	22	AAO06370	Human polypeptide
817	5	2.6	25	21	AAE09166	Hepatitis GB virus	890	5	2.6	37	22	AAO07512	Human polypeptide
818	5	2.6	25	23	ABB76462	Bovine IgA inducin	891	5	2.6	37	22	AAO13714	Human polypeptide
819	5	2.6	25	23	ABG62203	Eubacterial DNA Po	892	5	2.6	37	22	AAO13714	Conjugated peptide
820	5	2.6	25	23	ABG62247	Eubacterial DNA Po	893	5	2.6	37	22	AAO13714	Conjugated peptide
821	5	2.6	26	15	AAE52512	Human heavy chain	894	5	2.6	37	22	AAO13714	C glutamic prote
822	5	2.6	26	20	AAV07846	Human secreted pro	895	5	2.6	37	22	AAO13714	Immunomodulatory p
823	5	2.6	26	22	AAE59696	Rat SNS2 ion chann	896	5	2.6	37	22	AAO13714	Homoeodomain transc
824	5	2.6	27	23	AAU89448	Insulin/insulin-II	897	5	2.6	38	21	AAE22897	Rat ubiquitin prot
825	5	2.6	28	19	AAV21523	Human ubiquitin B2	898	5	2.6	38	21	AAO80350	Amino acid sequenc
826	5	2.6	28	20	AAV36437	Fragment of human	899	5	2.6	38	21	AAO80350	E. coli seryl-tRNA
827	5	2.6	28	21	AAV56397	Mutant ubiquitin-B	900	5	2.6	38	21	AAO80350	Arabidopsis thalia
828	5	2.6	29	18	AAW10846	MAB anti-HBsAg bin	901	5	2.6	38	21	AAO80350	Arabidopsis thalia
829	5	2.6	29	19	AAV21524	Human ubiquitin B	902	5	2.6	38	21	AAO80350	Arabidopsis thalia
830	5	2.6	29	20	AAV27862	Human secreted pro	903	5	2.6	38	22	ABB30388	Peptide #3039 enco
831	5	2.6	29	22	ABB37971	Peptide #5477 enco	904	5	2.6	38	22	ABB35561	Peptide #3067 enco
832	5	2.6	29	22	ABB23215	Protein #5214 enco	905	5	2.6	38	22	ABB39688	Peptide #7194 enco
833	5	2.6	29	22	AAW58602	Human brain expres	906	5	2.6	38	22	ABB20989	Protein #2988 enco
834	5	2.6	29	22	AAW71106	Human bone marrow	907	5	2.6	38	22	AAW56376	Human brain expres
835	5	2.6	29	22	AAW18851	Peptide #5285 enco	908	5	2.6	38	22	AAW60406	Human brain expres
836	5	2.6	29	22	AAW31382	Peptide #5419 enco	909	5	2.6	38	22	AAW68756	Human bone marrow
837	5	2.6	29	22	AAV97753	G. oxydans cytochr	910	5	2.6	38	22	AAW73042	Human bone marrow
838	5	2.6	29	23	ABG40900	Human peptide enco	911	5	2.6	38	22	AAW16579	Peptide #3013 enco
839	5	2.6	30	21	AAV96439	Variant human haem	912	5	2.6	38	22	AAW29062	Peptide #3099 enco
840	5	2.6	30	21	AAV56502	Ubiquitin-B amino	913	5	2.6	38	22	AAW33264	Peptide #7301 enco
841	5	2.6	30	22	ABB37821	Peptide #5327 enco	914	5	2.6	38	22	AAW04392	Peptide #2974 enco
842	5	2.6	30	22	ABB23090	Protein #5089 enco	915	5	2.6	38	22	ABB38338	Human peptide enco
843	5	2.6	30	22	AAW58437	Human brain expres	916	5	2.6	38	23	ABG42886	Human peptide enco
844	5	2.6	30	22	AAW70929	Human bone marrow	917	5	2.6	39	20	AAV11841	Human 5' EST secre
845	5	2.6	30	22	AAW88486	Human immune/haema	918	5	2.6	39	21	AAO58986	Arabidopsis thalia
846	5	2.6	30	22	AAW18734	Peptide #5168 enco	919	5	2.6	39	21	AAO58986	Arabidopsis thalia
847	5	2.6	30	22	AAW31219	Peptide #5256 enco	920	5	2.6	39	21	AAV66087	BRCAL-associated R
848	5	2.6	30	23	ABG40725	Human peptide enco	921	5	2.6	39	22	ABB40934	Peptide #8440 enco
849	5	2.6	31	19	AAV79678	Synthetic CS198 de	922	5	2.6	39	22	ABB25055	Protein #7054 enco
850	5	2.6	31	22	AAE07321	Human CS 198 pepti	923	5	2.6	39	22	AAW61796	Human brain expres
851	5	2.6	31	22	AAW83189	Human immune/haema	924	5	2.6	39	22	AAW74595	Human bone marrow
852	5	2.6	31	23	ABW44771	Human protective s	925	5	2.6	39	22	AAO7525	Human polypeptide
853	5	2.6	32	16	AAE85036	Peptide rI from th	926	5	2.6	39	22	AAO7525	Peptide #6789 enco
854	5	2.6	32	21	AAE44641	Human secreted pro	927	5	2.6	39	22	AAW34707	Peptide #8744 enco
855	5	2.6	32	21	AAE87774	Interleukin-10 pro	928	5	2.6	39	23	ABG44433	Human peptide enco
856	5	2.6	32	21	AAE87775	Interleukin-10 pro	929	5	2.6	40	13	AAE25790	[Arg16,21,22,25,27
857	5	2.6	32	22	ABB38466	Peptide #5972 enco	930	5	2.6	40	22	ABW1608	Peptide #9114 enco
858	5	2.6	32	22	AAW59077	Human brain expres	931	5	2.6	40	22	ABW25434	Protein #7433 enco
859	5	2.6	32	22	AAW71610	Human bone marrow	932	5	2.6	40	22	AAW62479	Human brain expres
860	5	2.6	32	22	AAW31907	Peptide #5944 enco	933	5	2.6	40	22	AAW75288	Human bone marrow
861	5	2.6	32	23	ABG411422	Human peptide enco	934	5	2.6	40	22	AAW84650	Human immune/haema
862	5	2.6	32	23	ABP26729	Streptococcus poly	935	5	2.6	40	22	AAW20559	Peptide #6993 enco
863	5	2.6	32	23	AAU70404	Human light chain	936	5	2.6	40	22	AAW35401	Peptide #9438 enco
864	5	2.6	33	18	AAW10861	MAB anti-HBsAg bin	937	5	2.6	40	23	ABG44928	Human peptide enco
865	5	2.6	33	22	ABW40890	Peptide #8396 enco	938	5	2.6	41	21	AAW34352	Gene 3 human secre
866	5	2.6	33	22	AAU27374	Novel bone marrow	939	5	2.6	41	22	AAO07432	Human polypeptide
867	5	2.6	33	22	AAW61750	Human brain expres	940	5	2.6	41	22	AAO09925	Human polypeptide
868	5	2.6	33	22	AAW74546	Human bone marrow	941	5	2.6	41	22	AAO13323	Human polypeptide
869	5	2.6	33	22	AAW34658	Peptide #8695 enco	942	5	2.6	41	22	AAW33596	Peptide #7633 enco
870	5	2.6	33	22	AAW47390	Conjugated peptide	943	5	2.6	41	22	ABG77086	Human colon cancer
871	5	2.6	33	22	AAW47391	Conjugated peptide	944	5	2.6	41	23	ABG43250	Human peptide enco
872	5	2.6	34	18	AAW19081	Trypanosoma cruzi	945	5	2.6	42	21	AAO11798	Arabidopsis thalia
873	5	2.6	34	20	AAW96175	IKK-alpha polypept	946	5	2.6	42	21	AAW56941	Arabidopsis thalia
874	5	2.6	34	21	AAW26476	Synthetic peptide	947	5	2.6	42	22	AAW64228	Human brain expres
875	5	2.6	34	21	AAW55417	Arabidopsis thalia	948	5	2.6	43	21	AAO11797	Arabidopsis thalia
876	5	2.6	34	22	AAW47416	Peptide #46 for il	949	5	2.6	43	21	AAO35517	Arabidopsis thalia
877	5	2.6	34	22	AAW92242	Human digestive sy	950	5	2.6	43	22	ABB15364	Human nervous syst
878	5	2.6	34	22	AAO02683	Human polypeptide	951	5	2.6	43	22	AAW86434	Human immune/haema
879	5	2.6	35	18	AAW15127	Phytase N-terminal	952	5	2.6	43	22	AAO07515	Human polypeptide
880	5	2.6	35	21	AAW57169	Human prostate can	953	5	2.6	43	22	AAO07612	Human polypeptide
881	5	2.6	35	21	AAW59361	Arabidopsis thalia	954	5	2.6	43	23	ABG65617	Human breast speci
882	5	2.6	35	22	AAW47388	Conjugated peptide	955	5	2.6	44	15	AAE54012	Characteristic pro
883	5	2.6	35	22	AAW47389	Conjugated peptide	956	5	2.6	44	21	AAE52537	Helicobacter pylor
884	5	2.6	35	22	AAW47392	Conjugated peptide	957	5	2.6	44	22	ABW96312	Human testicular a
885	5	2.6	36	21	AAW19327	Amino acid sequenc	958	5	2.6	44	22	ABG23704	Novel human diagn
886	5	2.6	36	22	ABB03530	Human musculoskele	959	5	2.6	44	22	ABB30222	Peptide #2873 enco

668	5	2.6	9	23	ABB94488	CTL epitope HLA pe	741	15	23	AAU86188	Oestradiol activat		
669	5	2.6	9	23	ABB94917	CTL epitope HLA pe	742	15	23	ABB04468	Human transcripti		
670	5	2.6	10	16	AAW21559	Corticotropin rele	743	16	19	AAW84025	Human CYP3A4 speci		
671	5	2.6	10	18	AAW15482	Human p100 protein	744	5	2.6	16	20	AAW97301	Peptide which indu
672	5	2.6	10	19	AAW44362	Breast cancer-asso	745	5	2.6	16	21	AAW65612	Oestrogen receptor
673	5	2.6	10	21	AAW27127	Human CASB619 prot	746	5	2.6	16	22	AAU08975	Human CDC25A inhib
674	5	2.6	10	22	AGG94052	Human complementar	747	5	2.6	16	23	ABB88202	C generalis mu-con
675	5	2.6	10	22	AGG94054	Human complementar	748	5	2.6	16	23	ABB88204	C generalis mu-con
676	5	2.6	10	22	AGG96450	Human complementar	749	5	2.6	16	23	ABB88427	C generalis mu-con
677	5	2.6	10	22	AGG84350	Arabidopsis thalia	750	5	2.6	16	23	ABB88428	C generalis mu-con
678	5	2.6	10	22	AGG86157	Saccharomyces cere	751	5	2.6	16	23	ABB88429	C generalis mu-con
679	5	2.6	10	22	AGG88110	Saccharomyces cere	752	5	2.6	16	23	AAU96349	Oestrogen receptor
680	5	2.6	10	22	AGG88111	Saccharomyces cere	753	5	2.6	16	23	ABB74556	DNA repair protein
681	5	2.6	10	22	AAU03225	Fruit fly G protei	754	5	2.6	17	19	AAW80775	Peptide that mimic
682	5	2.6	10	22	AAU03345	Fruit fly G protei	755	5	2.6	17	19	AAW62747	Streptococcus pneu
683	5	2.6	10	22	AAU03872	G protein-coupled	756	5	2.6	17	21	AAW56398	Mutant ubiquitin-B
684	5	2.6	10	22	AAW98714	Human breast cance	757	5	2.6	17	23	ABB09089	Human FAS partial
685	5	2.6	10	22	AAW30665	Neuropeptide F (NP	758	5	2.6	18	15	AAW47004	Ig kappa chain pos
686	5	2.6	10	23	AAW50005	Human D40 associat	759	5	2.6	18	19	AAW84020	Human CYP3A4 speci
687	5	2.6	10	23	ABB94564	CTL epitope HLA pe	760	5	2.6	18	20	AAW80885	D. melanogaster SL
688	5	2.6	10	23	ABB94728	CTL epitope HLA pe	761	5	2.6	18	20	AAW80894	D. melanogaster SL
689	5	2.6	10	23	ABB94856	CTL epitope HLA pe	762	5	2.6	19	19	AAW84024	Human CYP3A4 speci
690	5	2.6	10	23	ABB94962	CTL epitope HLA pe	763	5	2.6	19	21	AAW09271	Hepatitis GB virus
691	5	2.6	11	13	AAW28088	Cell-to-cell bindi	764	5	2.6	19	21	AAW98310	Alpha D peptide de
692	5	2.6	11	18	AAW11809	Fas ligand antigen	765	5	2.6	19	21	AAW98311	Conjugated peptide
693	5	2.6	11	19	AAW84023	Human CYP3A4 speci	766	5	2.6	19	22	AAW47366	Peptide derived fr
694	5	2.6	11	19	AAW84028	Human CYP3A4 speci	767	5	2.6	19	22	AAW84084	Mouse myosin heavy
695	5	2.6	11	22	AAW65972	Rat SNS1 ion chann	768	5	2.6	19	23	ABB77097	C-terminus of ubiq
696	5	2.6	12	13	AAW20711	Beta-2 integrin pe	769	5	2.6	19	23	AAW47907	Antigenic peptide
697	5	2.6	12	15	AAW67109	Antifertility pept	770	5	2.6	19	23	AAU82060	Cytotoxic T lympho
698	5	2.6	12	17	AAW02059	Human beta2 integr	771	5	2.6	20	16	AAW68766	T-cell stimulator
699	5	2.6	12	19	AAW84027	Human CYP3A4 speci	772	5	2.6	20	18	AAW35442	Alpha D peptide de
700	5	2.6	12	21	AAW98441	Human D peptide de	773	5	2.6	20	18	AAW11742	Peptide #10416 enc
701	5	2.6	12	23	AAU10818	Human Type I GnRH-	774	5	2.6	20	21	AAW98443	Human albumin-bind
702	5	2.6	12	23	ABB74483	DNA repair protein	775	5	2.6	20	22	ABB42910	Protein #8163 enco
703	5	2.6	13	17	AAW02560	Calcineurin bindin	776	5	2.6	20	22	ABB45302	Human brain expres
704	5	2.6	13	21	AAW27629	Human secreted pro	777	5	2.6	20	22	ABB26164	Human bone marrow
705	5	2.6	13	21	AAW14896	Human calcineurin-	778	5	2.6	20	22	AAW63811	Peptide #10767 enc
706	5	2.6	13	21	AAW98442	Alpha D peptide de	779	5	2.6	20	22	AAW76625	Human osteopontin
707	5	2.6	14	20	AAW00930	Prostate-tumour de	780	5	2.6	20	23	AAW36730	Insulin/insulin-li
708	5	2.6	14	21	AAW66097	BRCA1-associated R	781	5	2.6	20	23	ABG45852	T cell receptor pe
709	5	2.6	14	22	AAW65973	Rat SNS1 ion chann	782	5	2.6	21	18	AAW11765	T-cell receptor pe
710	5	2.6	14	22	AAW98866	Human peptide #141	783	5	2.6	21	18	AAW11743	T-cell receptor pe
711	5	2.6	14	22	AAW96940	Human peptide #215	784	5	2.6	21	18	AAW11752	T-cell receptor pe
712	5	2.6	14	22	AAW97433	Human peptide #708	785	5	2.6	21	19	AAW84012	Human CYP3A4 speci
713	5	2.6	14	22	AAW97744	Human peptide #101	786	5	2.6	21	22	ABB32457	Peptide #5108 enco
714	5	2.6	14	22	AAW63136	Peptide from B4-5	787	5	2.6	21	22	AAW20030	zinc finger helica
715	5	2.6	14	23	ABB76045	Human osteopontin	788	5	2.6	21	23	ABB76044	Human osteopontin
716	5	2.6	15	15	AAW47005	Ig kappa chain pos	789	5	2.6	21	23	AAU89162	Insulin/insulin-li
717	5	2.6	15	16	AAW81874	Chloroperoxidase t	790	5	2.6	21	23	AAU72609	T cell receptor im
718	5	2.6	15	16	AAW72034	Biologically activ	791	5	2.6	21	23	AAU72610	T cell receptor im
719	5	2.6	15	17	AAW02562	Calcineurin bindin	792	5	2.6	21	23	AAU72626	T cell receptor im
720	5	2.6	15	17	AAW02561	Calcineurin bindin	793	5	2.6	21	23	AAU72627	T cell receptor im
721	5	2.6	15	18	AAW11758	T-cell receptor pe	794	5	2.6	21	23	AAU72630	T cell receptor im
722	5	2.6	15	18	AAW11759	T-cell receptor pe	795	5	2.6	21	23	AAU72709	T cell receptor im
723	5	2.6	15	19	AAW20433	Human microtubule	796	5	2.6	22	19	AAW84013	Human CYP3A4 speci
724	5	2.6	15	19	AAW84021	Human CYP3A4 speci	797	5	2.6	22	20	AAW27485	E. coli beta-subun
725	5	2.6	15	21	AAW13482	C. pneumoniae Cp-S	798	5	2.6	22	20	AAW29739	Hepatitis B virus
726	5	2.6	15	21	AAW14897	Calcineurin-bindin	799	5	2.6	22	22	ABG04489	Novel human diagno
727	5	2.6	15	21	AAW14898	Calcineurin-bindin	800	5	2.6	22	23	AAU93938	Human cytochrome P
728	5	2.6	15	21	AAW13726	C-terminal peptide	801	5	2.6	23	21	AAW11419	A. thaliana SRP30
729	5	2.6	15	21	AAW66098	BRCA1-associated R	802	5	2.6	23	23	ABB44772	Human protective s
730	5	2.6	15	21	AAW65464	Estradiol activate	803	5	2.6	24	13	AAW24931	Peptide #2 used to
731	5	2.6	15	22	AAW99415	Vaccine related MH	804	5	2.6	24	20	AAW27489	E. coli beta-subun
732	5	2.6	15	22	AAW83158	Chlamydia Cp-Swib	805	5	2.6	24	20	AAW25839	Human secreted pro
733	5	2.6	15	22	AAW64800	Human zinc finger	806	5	2.6	24	22	ABB39492	Peptide #6998 enco
734	5	2.6	15	23	ABG67791	Human ADPI tryptic	807	5	2.6	24	22	AAW60183	Human brain expres
735	5	2.6	15	23	ABG67986	Human ADPI tryptic	808	5	2.6	24	22	AAW72801	Human bone marrow
736	5	2.6	15	23	ABB81761	Human alpha helica	809	5	2.6	24	22	AAG99705	ERA binding domain
737	5	2.6	15	23	ABG62301	Eubacterial DNA po	810	5	2.6	24	22	AAW88842	Human interleukin-
738	5	2.6	15	23	ABB81216	Human glutaryl-CoA	811	5	2.6	24	22	AAW88843	Human interleukin-
739	5	2.6	15	23	ABB08849	Human non-insulin	812	5	2.6	24	23	AAW23696	Fluorescently labe
740	5	2.6	15	23	ABB94129	Chlamydia peptide	813	5	2.6	24	23	ABB76465	Human IgA inducing

522	6	3.1	737	22	ABG10283	Novel human diagno	595	6	3.1	1020	22	ABBS58971	Drosophila melanog
523	6	3.1	747	23	ABBS0996	Herbicidally activ	596	6	3.1	1029	14	AAR38862	GC-A Rattus ratt
524	6	3.1	752	20	AAW97799	Streptococcus pneu	597	6	3.1	1033	23	ABBS4120	Drosophila melanog
525	6	3.1	752	20	AAW74407	C. trachomatis gid	598	6	3.1	1033	23	AAU82717	Amino acid sequenc
526	6	3.1	752	22	AAAB4212	Amino acid sequenc	599	6	3.1	1035	22	ABBS6062	Drosophila melanog
527	6	3.1	752	22	AAAB4213	Amino acid sequenc	600	6	3.1	1036	23	ABBS6062	Listeria monocytog
528	6	3.1	752	23	AAAG79339	Glycogen phosphory	601	6	3.1	1060	22	AAE09313	Sso/Tag DNA polyme
529	6	3.1	752	23	AAE24596	Human SR-cyp prote	602	6	3.1	1061	23	AAU11280	Human atrionatriur
530	6	3.1	755	22	AAAG90181	C. glutamicum prote	603	6	3.1	1061	23	AAU11281	Human atrionatriur
531	6	3.1	757	22	AAAG39928	Human polypeptide	604	6	3.1	1071	21	AAG36303	Arabidopsis thalia
532	6	3.1	759	22	ABBS68975	Drosophila melanog	605	6	3.1	1071	23	ABP27560	Streptococcus poly
533	6	3.1	759	23	AAAG70737	Human LP protein L	606	6	3.1	1072	21	AAG36302	Arabidopsis thalia
534	6	3.1	761	21	AAAG20796	Arabidopsis thalia	607	6	3.1	1075	20	AAV35498	C. pneumoniae prot
535	6	3.1	764	23	AAU81523	Human Che-2 protei	608	6	3.1	1075	22	ABBS11783	Human ANP-A recept
536	6	3.1	765	21	AAAG54100	Arabidopsis thalia	609	6	3.1	1098	22	ABG20355	Novel human diagno
537	6	3.1	773	23	ABBS1850	Herbicidally activ	610	6	3.1	1119	22	ABBS7825	Thermus aquaticus
538	6	3.1	781	22	AAU28171	Novel human secret	611	6	3.1	1119	23	AAE14716	Thermus aquaticus
539	6	3.1	784	20	AAW87760	Soybean lysine ket	612	6	3.1	1157	22	ABBS3351	Drosophila melanog
540	6	3.1	787	22	ABBS58149	Drosophila melanog	613	6	3.1	1187	16	AAR66451	AF-4 protein (enco
541	6	3.1	787	22	ABBS13669	Novel human diagno	614	6	3.1	1203	22	AAW79264	Human protein SEQ
542	6	3.1	787	23	AAE14673	Human phosphatidyl	615	6	3.1	1210	16	AAAG66450	AF-4 protein (enco
543	6	3.1	787	23	AAE14673	Human PRO protein,	616	6	3.1	1217	22	AAAG66450	AF-4 protein (enco
544	6	3.1	795	21	AAW70962	Rat Ras signalling	617	6	3.1	1217	22	AAAG66450	AF-4 protein (enco
545	6	3.1	797	21	AAW70963	Human Ras signalli	618	6	3.1	1305	22	ABBS7631	Drosophila melanog
546	6	3.1	805	22	ABBS59621	Drosophila melanog	619	6	3.1	1322	22	ABBS60275	Drosophila melanog
547	6	3.1	810	22	AAAG20795	Arabidopsis thalia	620	6	3.1	1327	22	ABBS58539	Drosophila melanog
548	6	3.1	812	22	AAAG60495	Human cell cycle a	621	6	3.1	1338	17	AAW91247	VIP2A(a)-VIP1A(a)
549	6	3.1	815	22	AAAG54099	Arabidopsis thalia	622	6	3.1	1338	18	AAW19520	Maize optimised-B
550	6	3.1	824	22	AAAG38907	E. coli growth and	623	6	3.1	1338	19	AAW46731	VIP2A(a)/VIP1A(a)
551	6	3.1	826	21	AAAG5050	Candida albicans p	624	6	3.1	1343	18	AAW31866	Mouse metastasis-a
552	6	3.1	826	23	ABBS1392	Herbicidally activ	625	6	3.1	1346	17	AAR91245	VIP2A(a) and VIP1A
553	6	3.1	827	22	AAU35764	Helicobacter pylor	626	6	3.1	1346	19	AAW19513	B. cereus VIP1A(a)
554	6	3.1	829	17	AAAG6578	Autotaxin derived	627	6	3.1	1363	22	ABBS64266	VIP1A(a)/VIP2A(a)
555	6	3.1	831	19	AAW54372	Carboxydothermus h	628	6	3.1	1443	22	ABBS25528	Drosophila melanog
556	6	3.1	844	22	ABBS3132	Drosophila melanog	629	6	3.1	1464	22	ABBS5043	Novel human diagno
557	6	3.1	850	23	AAW47574	Herbicidally cell cy	630	6	3.1	1477	16	AAR67691	S. cerevisiae scau
558	6	3.1	852	23	ABBS2825	Herbicidally activ	631	6	3.1	1477	18	AAW10424	Saccharomyces cere
559	6	3.1	856	13	AAAG23170	Mutant thermotabl	632	6	3.1	1477	20	AAW06819	Fumonisin-resistan
560	6	3.1	868	20	AAW00935	A. degensii DNA po	633	6	3.1	1543	22	ABBS69373	Drosophila melanog
561	6	3.1	868	23	AAE22112	Ammonifex degensii	634	6	3.1	1547	22	ABBS58643	Drosophila melanog
562	6	3.1	869	22	ABBS2777	Escherichia coli p	635	6	3.1	1572	23	ABBS7562	Novel human protei
563	6	3.1	882	15	AAAG3996	Thermophilic DNA p	636	6	3.1	1577	22	ABBS59387	Drosophila melanog
564	6	3.1	882	15	AAAG5749	Alpha-DNA polymera	637	6	3.1	1605	22	ABBS70375	Drosophila melanog
565	6	3.1	892	13	AAAG23169	Mutant thermotabl	638	6	3.1	1850	22	ABBS72210	Modified chicken v
566	6	3.1	892	13	AAAG23170	Taf DNA polymerase	639	6	3.1	1898	20	AAV30795	A human trichohyal
567	6	3.1	895	21	AAAG24233	Arabidopsis thalia	640	6	3.1	1920	22	ABBS64441	Drosophila melanog
568	6	3.1	903	20	AAW28713	Detargeted yeast m	641	6	3.1	2042	23	AAE24139	Human kinase (PKIN
569	6	3.1	904	22	ABBS68734	Drosophila melanog	642	6	3.1	2052	21	AAAB08634	A murine phosphati
570	6	3.1	906	20	AAW28702	Yeast multifunctio	643	6	3.1	2183	22	AAG84988	Shrimp white spot
571	6	3.1	906	20	AAW28712	Mutant yeast multi	644	6	3.1	2387	22	AAU01183	Rat glutamate tran
572	6	3.1	906	22	AAAB20180	Candida tropicalis	645	6	3.1	2390	20	AAV05494	BetaII spectrin p
573	6	3.1	906	22	AAAB20181	C. tropicalis multi	646	6	3.1	2400	22	ABBS20278	Novel human diagno
574	6	3.1	906	22	AAAB20182	C. tropicalis mult	647	6	3.1	2415	22	ABBS20279	Novel human diagno
575	6	3.1	906	22	AAAB20183	C. tropicalis mult	648	6	3.1	2424	22	ABBS58924	Drosophila melanog
576	6	3.1	915	17	AAAB6596	A2058 autotaxin pr	649	6	3.1	2451	22	ABBS71574	Drosophila melanog
577	6	3.1	915	22	AAW71987	Human melanoma aut	650	6	3.1	2618	22	ABBS02135	Novel human diagno
578	6	3.1	921	22	AAAG91905	C. glutamicum prote	651	6	3.1	2622	22	ABBS06418	Novel human diagno
579	6	3.1	925	23	ABBS91769	Herbicidally activ	652	6	3.1	2681	22	ABBS29184	Novel human diagno
580	6	3.1	931	22	ABBS23862	Novel human diagno	653	6	3.1	2803	23	ABBS08161	Human cytoskeleton
581	6	3.1	938	22	ABBS0562	Drosophila melanog	654	6	3.1	2958	22	ABBS71368	Drosophila melanog
582	6	3.1	950	22	AAU36368	Pseudomonas aerugi	655	6	3.1	3256	21	AAV50976	Human cell cycle p
583	6	3.1	956	21	AAAG42432	Arabidopsis thalia	656	6	3.1	3309	22	ABBS64158	Drosophila melanog
584	6	3.1	959	22	ABBS65368	Drosophila melanog	657	6	3.1	5533	22	ABBS65772	Drosophila melanog
585	6	3.1	960	22	ABBS58434	Drosophila melanog	658	6	3.1	5560	22	ABBS71160	Drosophila melanog
586	6	3.1	970	22	ABBS58624	Drosophila melanog	659	6	3.1	7107	22	ABBS58144	Drosophila melanog
587	6	3.1	982	22	ABBS62346	Drosophila melanog	660	6	2.6	5	20	AAW98057	Streptococcus agal
588	6	3.1	983	22	AAW74684	Human protease and	661	6	2.6	5	21	AAAB11068	S. agalactiae sorta
589	6	3.1	985	21	AAG36304	Arabidopsis thalia	662	6	2.6	5	23	AAW52885	Plasmodium falci
590	6	3.1	987	22	AAAB4633	Amino acid sequenc	663	6	2.6	7	22	ABBS1561	Zif268 zinc finger
591	6	3.1	1002	21	AAG42431	Arabidopsis thalia	664	6	2.6	7	23	ABBS74582	Transcription fact
592	6	3.1	1013	22	AAW84687	Amino acid sequenc	665	6	2.6	7	23	ABBS74701	Transcription fact
593	6	3.1	1013	22	AAE11053	Ashbya gossypii Gr	666	6	2.6	8	23	ABBS74583	Transcription fact
594	6	3.1	1014	23	ABBS08001	Human lipid metabo	667	6	2.6	9	21	ABBS26189	Human CASB619 prot



376	6	3.1	438	21	AAB03421	Soybean putative c	449	6	3.1	556	22	ABB59543	Drosophila melanog
377	6	3.1	439	21	AAV53797	Amino acid sequenc	450	6	3.1	558	21	AAB23175	Chicory germacrene
378	6	3.1	444	21	AAV19098	Protein encoded by	451	6	3.1	558	22	ABB64967	Drosophila melanog
379	6	3.1	444	23	ABP27271	Streptococcus poly	452	6	3.1	558	23	ABP27849	Streptococcus poly
380	6	3.1	444	23	ABB91258	Herbicidally activ	453	6	3.1	560	22	AAU22961	Novel human enzyme
381	6	3.1	444	23	ABB91259	Herbicidally activ	454	6	3.1	568	22	ABB64744	Drosophila melanog
382	6	3.1	446	19	AAW60232	Bacillus thuringie	455	6	3.1	568	22	AAE10823	Human gene 2 encod
383	6	3.1	446	21	AAV59279	WAR toxin from B.	456	6	3.1	571	22	AAV89979	Thrombosucin, Uni
384	6	3.1	449	17	AAV91249	Vip2A(a) protein w	457	6	3.1	571	23	AAO17248	A thaliana RKSII su
385	6	3.1	449	18	AAV19519	Maize optimised-B.	458	6	3.1	576	22	ABB62520	Drosophila melanog
386	6	3.1	449	19	AAW46730	Vip2A(a) with a va	459	6	3.1	578	22	ABB59076	Drosophila melanog
387	6	3.1	452	21	AAAG32061	Arabidopsis thalia	460	6	3.1	579	23	AAO17257	A thaliana RKSIII
388	6	3.1	454	21	AAAB41897	Human OREF ORF1661	461	6	3.1	579	23	AAO17260	A thaliana recepto
389	6	3.1	456	21	AAAG51360	Arabidopsis thalia	462	6	3.1	580	23	AAO17311	A thaliana recepto
390	6	3.1	456	18	AAW14592	Streptococcus pneu	463	6	3.1	583	22	AAU30814	Novel human secret
391	6	3.1	462	17	AAV91238	B. cereus Vip2A(a)	464	6	3.1	588	22	AAO17251	A thaliana RKSII s
392	6	3.1	462	18	AAV19508	B. cereus Vip2A(a)	465	6	3.1	591	23	AAO17256	A thaliana RKSIII
393	6	3.1	462	19	AAW46710	Vegetative insecti	466	6	3.1	592	23	AAO17258	A thaliana RKSIII
394	6	3.1	463	22	ABB64109	Drosophila melanog	467	6	3.1	596	23	AAO17271	A thaliana recepto
395	6	3.1	466	22	AAV95125	Human protein sequ	468	6	3.1	597	23	AAO17259	A thaliana recepto
396	6	3.1	466	23	AAU76040	Human ADAM-H9 prot	469	6	3.1	598	21	AAAG30868	Arabidopsis thalia
397	6	3.1	467	21	AAAG30869	Arabidopsis thalia	470	6	3.1	600	11	AAAG39922	M.hypopneumoniae HS
398	6	3.1	467	22	ABB50210	Human transcriptio	471	6	3.1	600	14	AAAR43003	Mycoplasma 74.5kd
399	6	3.1	472	20	AAV41722	Human PRO871 prote	472	6	3.1	600	23	AAU82971	S. cerevisiae TPB1
400	6	3.1	472	21	AAAB44278	Human PRO871 (UNQ4	473	6	3.1	603	16	AAAR74620	Human lung tumour
401	6	3.1	472	22	AAU23072	Human PRO polypept	474	6	3.1	607	20	AAV14149	DeltaChy DNA polym
402	6	3.1	472	22	AAAB6865	Human peptidyl-pro	475	6	3.1	607	20	AAV14300	DeltaChy DNA polym
403	6	3.1	472	23	ABB95466	Human angioogenesis	476	6	3.1	607	23	AAO17316	A thaliana recepto
404	6	3.1	472	23	ABB84860	Human PRO871 prote	477	6	3.1	610	21	AAAG36219	Arabidopsis thalia
405	6	3.1	473	21	AAV59829	Soybean histone de	478	6	3.1	612	21	AAAG51362	Arabidopsis thalia
406	6	3.1	475	14	AAAR41824	Methylobacillus g1	479	6	3.1	615	21	AAAG36218	Arabidopsis thalia
407	6	3.1	477	22	AAU69456	Human purified sec	480	6	3.1	615	21	AAAG51359	Arabidopsis thalia
408	6	3.1	482	23	ABP39464	Staphylococcus epi	481	6	3.1	617	23	ABB93952	Herbicidally activ
409	6	3.1	485	21	AAAG49236	Arabidopsis thalia	482	6	3.1	619	22	ABB66630	Drosophila melanog
410	6	3.1	485	19	AAE09718	Human ubiquitin ca	483	6	3.1	621	14	AAAR41230	Taurine transporte
411	6	3.1	494	22	AAW47017	Arabidopsis thalia	484	6	3.1	621	22	ABB64486	Drosophila melanog
412	6	3.1	495	21	AAAG51363	Arabidopsis thalia	485	6	3.1	621	22	ABB64485	Amino acid sequenc
413	6	3.1	501	22	AAU33796	Staphylococcus aur	486	6	3.1	622	22	AAAB94338	Human protein sequ
414	6	3.1	504	22	AAU336910	Staphylococcus aur	487	6	3.1	622	22	AAAB95098	Human protein sequ
415	6	3.1	507	16	AAW01499	70 kD cellobiohydr	488	6	3.1	622	23	AAU74330	Human cytoskeleton
416	6	3.1	508	22	AAE09309	Sulfobolus sulfata	489	6	3.1	624	23	AAE15257	Human RNA metaboli
417	6	3.1	512	22	ABB65259	Drosophila melanog	490	6	3.1	625	19	AAW47023	Arabidopsis thalia
418	6	3.1	518	22	AAU02059	Synthetic human mu	491	6	3.1	625	23	AAO17246	A thaliana recepto
419	6	3.1	520	21	AAAB19740	Arabidopsis acyl C	492	6	3.1	625	23	AAO17254	A thaliana RKSIII
420	6	3.1	520	21	AAV94512	Arabidopsis diacyl	493	6	3.1	625	23	ABB91554	Herbicidally activ
421	6	3.1	520	21	AAV94524	A. thaliana diacyl	494	6	3.1	628	21	AAAG30867	Arabidopsis thalia
422	6	3.1	520	21	AAAG32060	Arabidopsis thalia	495	6	3.1	628	22	AAW47174	Human polypeptide
423	6	3.1	520	21	AAAG35807	Arabidopsis thalia	496	6	3.1	628	23	AAO17255	A thaliana RKSIII
424	6	3.1	520	21	AAV96853	A. thaliana diacyl	497	6	3.1	628	23	AAO17269	A thaliana recepto
425	6	3.1	520	21	AAV54143	Acyl-CoA:cholester	498	6	3.1	628	23	ABB91241	Herbicidally activ
426	6	3.1	520	22	AAU00462	Arabidopsis thalia	499	6	3.1	630	21	AAAB25103	Pinus radiata cell
427	6	3.1	520	23	ABB91743	Herbicidally activ	500	6	3.1	632	23	ABB91401	Herbicidally activ
428	6	3.1	522	22	ABB63031	Drosophila melanog	501	6	3.1	632	23	ABB07638	Human speckled l10
429	6	3.1	522	22	ABB65580	Drosophila melanog	502	6	3.1	640	20	ABB61283	Drosophila melanog
430	6	3.1	523	21	AAAG31084	Arabidopsis thalia	503	6	3.1	650	20	AAV14962	Amino acid sequenc
431	6	3.1	523	23	ABB93214	Herbicidally activ	504	6	3.1	650	22	AAAB83262	C elegans FAIPA SE
432	6	3.1	524	21	AAAG32057	Arabidopsis thalia	505	6	3.1	650	22	AAAB83274	C elegans FAIPA SE
433	6	3.1	524	21	AAAG49211	Arabidopsis thalia	506	6	3.1	660	22	ABB67711	Drosophila melanog
434	6	3.1	524	23	ABB91744	Herbicidally activ	507	6	3.1	661	22	ABB60769	Drosophila melanog
435	6	3.1	527	23	ABB92687	Herbicidally activ	508	6	3.1	665	22	ABB64312	Drosophila melanog
436	6	3.1	528	22	AAV78918	Human protein sequ	509	6	3.1	679	22	ABG18804	Novel human diagno
437	6	3.1	529	23	ABP43478	Human secreted pro	510	6	3.1	680	22	ABB68821	Drosophila melanog
438	6	3.1	530	21	AAAG36220	Arabidopsis thalia	511	6	3.1	689	23	ABB07636	Human speckled l10
439	6	3.1	540	22	ABB11354	Human Ser/Arg rich	512	6	3.1	690	22	ABB63142	Drosophila melanog
440	6	3.1	542	22	AAAB94243	Human protein sequ	513	6	3.1	702	22	AAAB6724	S. solfatarius DN
441	6	3.1	542	22	AAAB95042	Human protein sequ	514	6	3.1	709	22	ABB6396	TSHR-GS-alpha fusi
442	6	3.1	542	23	ABB08186	Human nucleotide r	515	6	3.1	718	22	ABB58626	Drosophila melanog
443	6	3.1	542	23	ABB97456	Novel human protei	516	6	3.1	720	22	ABB58576	Drosophila melanog
444	6	3.1	542	23	AAU81996	Human secreted pro	517	6	3.1	722	12	AAAR10055	Cephalosporin C am
445	6	3.1	547	21	AAV96854	A. thaliana AS11 m	518	6	3.1	725	22	AAAB99036	Human somatostatin
446	6	3.1	549	22	ABB50168	Human transcriptio	519	6	3.1	727	20	AAV43523	Human CCCTC-bindin
447	6	3.1	552	23	ABB91328	Herbicidally activ	520	6	3.1	728	20	AAV43520	Chicken CCCTC-bind
448	6	3.1	553	19	AAW47013	Daucus carota SERK	521	6	3.1	736	23	ABB57361	Mouse ischaemic co

230	6	3.1	283	22	AAB95629	Human protein sequ	303	6	3.1	355	23	ABB09277	G protein-coupled
231	6	3.1	283	23	ABG61546	Human transporter	304	6	3.1	355	23	AAU79335	Human inhibitory G
232	6	3.1	284	22	AAB76732	Corynebacterium gl	305	6	3.1	359	22	ABB67482	Drosophila melanog
233	6	3.1	284	22	AAB76734	Corynebacterium gl	306	6	3.1	359	22	ABB52703	Escherichia coli p
234	6	3.1	286	20	AAI32172	Chlamydia trachoma	307	6	3.1	361	22	ABY22206	Biothym marker p
235	6	3.1	287	21	ABAB3685	Human protein frag	308	6	3.1	361	23	ABY54428	Lactococcus lactis
236	6	3.1	287	22	AAK39480	Human polypeptide	309	6	3.1	368	21	ABY32059	Arabidopsis thalia
237	6	3.1	290	21	ABG48998	Arabidopsis thalia	310	6	3.1	369	17	AAW06822	Turkey herpes viru
238	6	3.1	291	23	AAE15249	Human RNA metaboli	311	6	3.1	369	21	AAK34421	Arabidopsis thalia
239	6	3.1	294	11	AAK06435	Halooacetate dehalo	312	6	3.1	369	21	AAK37344	Arabidopsis thalia
240	6	3.1	294	21	AAK05510	Arabidopsis thalia	313	6	3.1	372	20	AAI07102	Colon cancer assoc
241	6	3.1	299	21	AAK49929	Arabidopsis thalia	314	6	3.1	375	22	AAW40033	Human polypeptide
242	6	3.1	299	22	AAK65613	Amino acid sequenc	315	6	3.1	378	22	AAK60643	Arabidopsis thalia
243	6	3.1	301	21	AAE50380	Human uncoupling p	316	6	3.1	380	22	AAK94780	Human protein sequ
244	6	3.1	301	23	AAE16767	Human transporter	317	6	3.1	383	19	AAW98461	H. pylori GHPO 718
245	6	3.1	303	21	AAK09218	Arabidopsis thalia	318	6	3.1	385	20	AAW78473	Human 68772 protei
246	6	3.1	305	20	AAI37455	Protein involved i	319	6	3.1	385	21	AAK41850	Human ORFX ORF1614
247	6	3.1	306	21	AAK54149	Human pancreatic c	320	6	3.1	387	22	ABY59239	Drosophila melanog
248	6	3.1	308	23	ABP30403	Streptococcus poly	321	6	3.1	391	21	AAK37343	Arabidopsis thalia
249	6	3.1	310	22	ABG04751	Novel human diagno	322	6	3.1	391	22	AAE12467	Tobacco translatio
250	6	3.1	312	21	AAK09217	Arabidopsis thalia	323	6	3.1	392	22	ABY58291	Drosophila melanog
251	6	3.1	312	21	AAK11045	Arabidopsis thalia	324	6	3.1	392	22	ABG09960	Novel human diagno
252	6	3.1	312	21	AAK40400	Arabidopsis thalia	325	6	3.1	393	21	AAK51361	Arabidopsis thalia
253	6	3.1	312	22	AAK71710	Human olfactory re	326	6	3.1	394	22	ABG12146	Novel human diagno
254	6	3.1	312	22	AAK72812	Human olfactory re	327	6	3.1	395	21	AAK49780	Arabidopsis thalia
255	6	3.1	312	22	AAK72978	Olfactory receptor	328	6	3.1	395	23	ABY93740	Herbicidally activ
256	6	3.1	312	23	ABY09988	Herbicidally activ	329	6	3.1	396	19	AAW80682	S. pneumoniae prot
257	6	3.1	314	22	ABY61195	Human testicular a	330	6	3.1	397	23	ABY22240	Herbicidally activ
258	6	3.1	314	22	AAK95512	Human reproductive	331	6	3.1	399	22	ABY63309	Drosophila melanog
259	6	3.1	314	23	ABY47431	Listeria monocytog	332	6	3.1	399	22	AAU17440	Novel signal trans
260	6	3.1	315	22	AAK72979	Olfactory receptor	333	6	3.1	401	23	ABY92241	Herbicidally activ
261	6	3.1	315	23	ABY6310	Human protein phos	334	6	3.1	402	22	ABY93220	Human protein sequ
262	6	3.1	315	23	ABY27294	Streptococcus poly	335	6	3.1	402	22	ABY95804	Human protein sequ
263	6	3.1	316	21	AAK56777	Human prostate can	336	6	3.1	402	23	ABY97296	Novel human protei
264	6	3.1	319	22	ABG06331	Novel human diagno	337	6	3.1	403	21	AAK05758	Arabidopsis thalia
265	6	3.1	321	21	AAK34423	Arabidopsis thalia	338	6	3.1	403	21	AAK49835	Arabidopsis thalia
266	6	3.1	321	21	AAK94272	Corn arginyl-tRNA	339	6	3.1	404	23	ABY91257	Herbicidally activ
267	6	3.1	322	21	AAK92821	C. pneumoniae CPN1	340	6	3.1	406	20	AAK49152	Amino acid sequenc
268	6	3.1	323	22	AAK39441	Human polypeptide	341	6	3.1	406	21	AAK05757	Arabidopsis thalia
269	6	3.1	324	22	ABY70087	Drosophila melanog	342	6	3.1	406	21	AAK49834	Arabidopsis thalia
270	6	3.1	325	21	AAK10582	Arabidopsis thalia	343	6	3.1	406	22	AAK98360	Escherichia coli p
271	6	3.1	325	23	ABY47439	Listeria monocytog	344	6	3.1	408	21	AAK37342	Arabidopsis thalia
272	6	3.1	326	21	AAK10581	Arabidopsis thalia	345	6	3.1	408	21	AAK77123	Human neurotransm
273	6	3.1	326	21	AAK49782	Arabidopsis thalia	346	6	3.1	409	21	AAK51364	Arabidopsis thalia
274	6	3.1	327	22	ABG04887	Novel human diagno	347	6	3.1	409	22	ABY93991	Human protein sequ
275	6	3.1	327	22	AAK72811	Human olfactory re	348	6	3.1	410	17	AAK91248	VIP2A(a) protein w
276	6	3.1	328	22	AAU51216	Propionibacterium	349	6	3.1	410	17	AAK91251	VIP2A(a) protein w
277	6	3.1	329	21	AAK27834	Arabidopsis thalia	350	6	3.1	410	18	AAW19517	Amino acid sequenc
278	6	3.1	330	21	AAK05509	Arabidopsis thalia	351	6	3.1	410	18	AAW19518	Ras binding protei
279	6	3.1	331	8	AAK70328	Biotin-synthase ge	352	6	3.1	410	19	AAK46728	Drosophila melanog
280	6	3.1	331	23	ABY91120	Herbicidally activ	353	6	3.1	410	19	AAK46729	Novel signal trans
281	6	3.1	332	21	AAK29966	Arabidopsis thalia	354	6	3.1	412	20	AAK49151	Human keratin 45.8
282	6	3.1	335	23	ABY09580	Chlamydia pneumoni	355	6	3.1	413	20	AAK05724	N-terminal region
283	6	3.1	336	22	ABY05292	S. epidermidis ope	356	6	3.1	413	22	ABY70347	N-terminal choline
284	6	3.1	337	9	AAK81191	Biotin synthetase	357	6	3.1	415	22	AAU17462	Amino acid sequenc
285	6	3.1	339	21	AAK10580	Arabidopsis thalia	358	6	3.1	417	23	AAK48190	Drosophila melanog
286	6	3.1	339	21	AAK49781	Arabidopsis thalia	359	6	3.1	419	20	AAK49243	Arabidopsis thalia
287	6	3.1	341	23	ABY7246	Mouse ischaemic co	360	6	3.1	419	20	AAK32182	N-terminal choline
288	6	3.1	342	21	AAK34422	Arabidopsis thalia	361	6	3.1	419	20	AAK49149	Amino acid sequenc
289	6	3.1	346	21	AAK32084	Arabidopsis thalia	362	6	3.1	420	21	AAK23364	Arabidopsis thalia
290	6	3.1	347	22	AAK41227	Human polypeptide	363	6	3.1	421	22	ABY64448	Drosophila melanog
291	6	3.1	352	21	AAK29965	Arabidopsis thalia	364	6	3.1	422	21	AAK05756	Arabidopsis thalia
292	6	3.1	352	23	ABY92713	Herbicidally activ	365	6	3.1	422	21	AAK49833	Arabidopsis thalia
293	6	3.1	352	23	AAU76041	Human ADAM-H9 prot	366	6	3.1	425	20	AAK49248	N-terminal region
294	6	3.1	353	23	ABY49274	Listeria monocytog	367	6	3.1	425	20	AAK32187	N-terminal choline
295	6	3.1	354	20	AAK07104	Colon cancer assoc	368	6	3.1	425	21	AAK68783	Amino acid sequenc
296	6	3.1	354	21	AAK85290	Human G-alpha-11 a	369	6	3.1	425	22	AAK79121	Human protein SEQ
297	6	3.1	354	22	AAK99064	Human G-protein al	370	6	3.1	427	23	ABY41163	Human ovarian anti
298	6	3.1	354	23	ABY09273	G protein-coupled	371	6	3.1	430	21	AAK23363	Arabidopsis thalia
299	6	3.1	355	21	AAK29964	Arabidopsis thalia	372	6	3.1	431	21	AAK32058	Arabidopsis thalia
300	6	3.1	355	21	AAK85149	Human G-alpha-12 a	373	6	3.1	432	22	ABY61904	Drosophila melanog
301	6	3.1	355	22	AAK99065	Human G-protein al	374	6	3.1	433	20	AAK49245	N-terminal region
302	6	3.1	355	23	ABY09274	G protein-coupled	375	6	3.1	433	20	AAK32184	N-terminal choline

84	6	3.1	69	21	AAG36767	Arabidopsis thalia	157	6	3.1	155	22	ABG27281	Novel human diagno
85	6	3.1	73	22	ABG24453	Novel human diagno	158	6	3.1	156	21	AAB57153	Human prostate can
86	6	3.1	73	22	AAU18639	Human lung antigen	159	6	3.1	156	22	ABG14998	Novel human diagno
87	6	3.1	73	22	AAM89495	Human immune/haema	160	6	3.1	157	21	AAB16272	Pinus radiata amyl
88	6	3.1	74	21	AAB16773	Bacteriophage Dp-1	161	6	3.1	157	21	AAB16272	Human protein sequ
89	6	3.1	76	21	AAB41435	Human ORFX ORF199	162	6	3.1	162	22	AAB95071	Streptococcus pneu
90	6	3.1	76	22	ABG05125	Novel human diagno	163	6	3.1	166	18	AAW14568	Arabidopsis thalia
91	6	3.1	78	22	AAG01097	Human secreted pro	164	6	3.1	166	22	AAB93936	Human protein sequ
92	6	3.1	79	22	AAG05346	Propionibacterium	165	6	3.1	173	21	AAG04363	Arabidopsis thalia
93	6	3.1	80	23	ABP3212	Human ORF2185 prot	166	6	3.1	173	23	AAU72498	Arabidopsis thalia
94	6	3.1	83	21	AAG54512	Zea mays protein f	167	6	3.1	179	20	AAU34846	Amino acid sequenc
95	6	3.1	85	22	ABG57535	Human testicular a	168	6	3.1	179	21	AAB25390	Pinus radiata cell
96	6	3.1	85	22	AAM95028	Human reproductive	169	6	3.1	180	23	ABW49350	Listeria monocytog
97	6	3.1	88	22	AAU22171	Human cardiovascu	170	6	3.1	184	18	AAW14589	Streptococcus pneu
98	6	3.1	89	22	AAO12560	Human polypeptide	171	6	3.1	186	23	ABG77118	Amino acid sequenc
99	6	3.1	90	21	AAU53798	Amino acids 145-23	172	6	3.1	193	22	ABG01292	Novel human diagno
100	6	3.1	96	22	AAM84545	Human immune/haema	173	6	3.1	196	19	AAW82630	Ehrlichia sp. HGE-
101	6	3.1	98	21	AAB53527	Human colon cancer	174	6	3.1	196	21	AAW82643	Ehrlichia sp. exte
102	6	3.1	105	20	AAG60007	Human endometrium	175	6	3.1	196	21	AAW78554	Ehrlichia antigen
103	6	3.1	105	21	AAG12168	Arabidopsis thalia	176	6	3.1	196	23	AAU73200	Human granulocytic
104	6	3.1	105	21	AAU90691	Human spectrin ple	177	6	3.1	196	23	AAU73213	Human granulocytic
105	6	3.1	105	22	AAU44839	Propionibacterium	178	6	3.1	198	22	ABG00682	Novel human diagno
106	6	3.1	107	23	ABG90676	Chlamydia pneumoni	179	6	3.1	198	22	ABG67520	Amino acid sequenc
107	6	3.1	107	23	AAW48199	Human zinc finger	180	6	3.1	198	23	ABP38614	Staphylococcus epi
108	6	3.1	109	22	ABG30376	Novel human diagno	181	6	3.1	204	22	ABW70670	Drosophila melanog
109	6	3.1	110	23	ABP30180	Streptococcus poly	182	6	3.1	206	22	ABW69257	Drosophila melanog
110	6	3.1	111	20	AAU37080	Chlamydia trachoma	183	6	3.1	212	18	AAW14588	Streptococcus pneu
111	6	3.1	111	23	ABP42966	Human ovarian anti	184	6	3.1	214	20	AAU35473	Chlamydia pneumoni
112	6	3.1	113	23	AAU66759	Propionibacterium	185	6	3.1	219	21	AAU49000	Arabidopsis thalia
113	6	3.1	113	22	ABG24902	Novel human diagno	186	6	3.1	220	23	ABP01204	Human ORFX protein
114	6	3.1	117	23	ABP46893	Streptococcus poly	187	6	3.1	221	21	AAW1753	Human ORFX ORF1517
115	6	3.1	118	22	AAM80051	Human protein sequ	188	6	3.1	221	21	AAU94283	Corn arginyl-tRNA
116	6	3.1	118	22	AAU02054	Synthetic human ta	189	6	3.1	223	21	AAG48999	Arabidopsis thalia
117	6	3.1	119	18	AAU23002	Canine herpesvirus	190	6	3.1	224	22	AAU94724	Human protein sequ
118	6	3.1	119	19	AAU72662	Canine herpes viru	191	6	3.1	227	20	AAU10844	Amino acid sequenc
119	6	3.1	119	22	ABU51319	Canine herpes viru	192	6	3.1	228	21	AAG49931	Arabidopsis thalia
120	6	3.1	120	22	AAU40633	Propionibacterium	193	6	3.1	230	23	ABW54441	Lactococcus lactis
121	6	3.1	126	22	ABW67055	Drosophila melanog	194	6	3.1	232	21	AAG49930	Arabidopsis thalia
122	6	3.1	127	23	ABW47947	Listeria monocytog	195	6	3.1	232	18	AAW14590	Streptococcus pneu
123	6	3.1	130	22	ABG23863	Novel human diagno	196	6	3.1	235	22	AAG90926	C glutamicum prote
124	6	3.1	132	21	AAW41950	Human ORFX ORF1714	197	6	3.1	236	22	AAG92550	C glutamicum prote
125	6	3.1	132	23	ABW55438	Lactococcus lactis	198	6	3.1	238	18	AAW14587	Streptococcus pneu
126	6	3.1	134	21	ABW58888	Breast and ovarian	199	6	3.1	238	22	ABG15228	Novel human diagno
127	6	3.1	134	21	AAG55969	Zea mays protein f	200	6	3.1	240	21	AAU94275	Wheat arginyl-tRNA
128	6	3.1	136	22	AAU92870	Human protein sequ	201	6	3.1	242	22	AAW79067	Human protein sequ
129	6	3.1	138	22	ABW27701	Human peptide #352	202	6	3.1	242	22	AAM39926	Human polypeptide
130	6	3.1	138	22	ABW32872	Peptide #378 encod	203	6	3.1	243	15	AAW63792	Bacillus cereus ve
131	6	3.1	138	22	ABW18354	Protein #353 encod	204	6	3.1	243	23	ABW47402	Listeria monocytog
132	6	3.1	138	22	AAW63674	Human brain expres	205	6	3.1	244	23	ABP39939	Staphylococcus epi
133	6	3.1	138	22	AAM66058	Human bone marrow	206	6	3.1	247	21	AAU74389	Neisseria gonorrhoe
134	6	3.1	138	22	AAU13927	Peptide #361 encod	207	6	3.1	254	23	ABW79353	Human ovary specif
135	6	3.1	138	22	AAM26333	Peptide #370 encod	208	6	3.1	254	23	ABW77351	Plodia interpuncte
136	6	3.1	138	22	AAU01670	Peptide #352 encod	209	6	3.1	255	22	AAW94259	Human protein sequ
137	6	3.1	138	23	ABW35706	Human peptide enco	210	6	3.1	258	21	AAU56510	Human Jurkat cell
138	6	3.1	139	18	AAW23003	Canine herpesvirus	211	6	3.1	259	22	AAM39028	Human polypeptide
139	6	3.1	139	19	AAW22663	Canine herpes viru	212	6	3.1	259	23	ABW97275	Novel human protei
140	6	3.1	139	22	AAW51320	Canine herpes viru	213	6	3.1	261	20	AAU37150	Chlamydia trachoma
141	6	3.1	140	21	AAB31354	Human ORFX ORF2918	214	6	3.1	261	22	AAW41712	Human polypeptide
142	6	3.1	141	22	AAW41819	Human polypeptide	215	6	3.1	261	22	AAB93180	Human protein sequ
143	6	3.1	142	21	AAG04365	Arabidopsis thalia	216	6	3.1	261	23	ABW53449	Lactococcus lactis
144	6	3.1	143	19	AAU96163	S. pneumoniae deri	217	6	3.1	262	22	ABW64085	Drosophila melanog
145	6	3.1	145	22	AAU87315	Novel central nerv	218	6	3.1	263	21	AAG11047	Arabidopsis thalia
146	6	3.1	146	22	AAU16613	Human novel secret	219	6	3.1	263	21	AAG40402	Arabidopsis thalia
147	6	3.1	147	22	ABG04065	Novel human diagno	220	6	3.1	265	21	AAU94451	Human inflammation
148	6	3.1	148	21	AAG04364	Arabidopsis thalia	221	6	3.1	266	21	AAG11046	Arabidopsis thalia
149	6	3.1	148	21	AAG14968	Arabidopsis thalia	222	6	3.1	266	21	AAG40401	Arabidopsis thalia
150	6	3.1	148	22	AAB96571	Putative P. abyssi	223	6	3.1	267	22	AAM40814	Human polypeptide
151	6	3.1	148	22	AAB66021	Human secreted pro	224	6	3.1	271	22	AAW41266	Human polypeptide
152	6	3.1	148	22	AAB66022	Human secreted pro	225	6	3.1	273	21	AAG27836	Arabidopsis thalia
153	6	3.1	149	18	AAW22993	Canine herpesvirus	226	6	3.1	276	21	AAG27835	Arabidopsis thalia
154	6	3.1	149	19	AAW72652	Canine herpes viru	227	6	3.1	276	21	AAG31086	Arabidopsis thalia
155	6	3.1	149	20	AAU34841	Chlamydia pneumoni	228	6	3.1	279	21	AAG31085	Arabidopsis thalia
156	6	3.1	149	22	AAB51309	Canine herpes viru	229	6	3.1	279	21	AAG32062	Arabidopsis thalia

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:54:48 : Search time 36 Seconds  
(without alignments)  
710.670 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 192

Sequence: 1 MRKEVTPMLNKNYPGPQF.....EEQDKEMTSKQHLFLVRKN 192

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

A\_Geneseq\_101002:\*

1:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	100.0	192	21	AA196805
2	185	96.4	192	23	ABP25486
3	12	6.2	176	22	AAU37821
4	12	6.2	22	AAU37974	Streptococcus pneu
5	12	6.2	214	23	ABP25487
6	11	5.7	91	23	ABP25067
7	8	4.2	251	23	ABP53482
8	8	4.2	303	23	ABP28656
9	8	4.2	733	20	AA130937
10	8	4.2	733	20	AA122202
					p90 ribosomal S6 k

11	7	3.6	15	20	AAW93973	Human IRS-1 and IR
12	7	3.6	129	22	AC82114	S. epidermidis ope
13	7	3.6	132	23	ABP40049	Staphylococcus epi
14	7	3.6	157	22	AB47229	Yeast Mfn. Saccha
15	7	3.6	159	20	AAW93976	Human IRS-1 and IR
16	7	3.6	235	22	AAW40165	Human polypeptide
17	7	3.6	240	21	AAW09871	Arabidopsis thalia
18	7	3.6	248	21	AAW09870	Arabidopsis thalia
19	7	3.6	254	22	AAW36598	Human FLEXHT-20 pr
20	7	3.6	256	21	AAW09869	Arabidopsis thalia
21	7	3.6	266	22	AAW030761	Novel human secret
22	7	3.6	280	21	AAW42139	Human ORFX ORF1903
23	7	3.6	283	22	AAW41951	Human polypeptide
24	7	3.6	284	20	AAW41211	H. influenzae Yig
25	7	3.6	317	19	AAW73507	Human ATG-1709 pro
26	7	3.6	317	19	AAW37816	Human secreted apo
27	7	3.6	324	22	AAW54873	Propionibacterium
28	7	3.6	334	21	AAW41766	Human ORFX ORF1530
29	7	3.6	376	22	AAW51364	Feline infectious
30	7	3.6	380	9	AAW82001	Sequence of phenyl
31	7	3.6	382	14	AAW39224	Nucleocapsid prote
32	7	3.6	458	23	ABW77059	Human protein sequ
33	7	3.6	628	23	ABW54651	Lactococcus lactis
34	7	3.6	782	23	AAW81501	Chlamydomonas inte
35	7	3.6	823	21	AAW42436	Arabidopsis thalia
36	7	3.6	849	21	AAW42816	Human ORFX ORF2580
37	7	3.6	1242	20	AAW13461	Amino acid sequenc
38	7	3.6	1242	20	AAW93972	Human IRS-1 and IR
39	7	3.6	1242	22	AAW83921	Amino acid sequenc
40	7	3.6	1243	16	AAW67708	Insulin receptor s
41	7	3.6	1321	17	AAW69994	Mouse IRS-2. Mus
42	7	3.6	1489	22	ABG19124	Novel human diagno
43	7	3.6	2386	18	AAW13153	S. pombe Rad3 poly
44	6	3.1	19	21	AAW98312	Alpha D peptide de
45	6	3.1	25	22	AAW30370	Novel human secret
46	6	3.1	29	22	ABG29861	Novel human diagno
47	6	3.1	38	15	AAW50672	G-protein coupled
48	6	3.1	38	17	AAW02864	G-protein coup-ed
49	6	3.1	38	22	ABW40724	Peptide #8230 enco
50	6	3.1	38	22	ABW24951	Protein #6950 enco
51	6	3.1	38	22	AAW61586	Human brain expres
52	6	3.1	38	22	AAW74377	Human bone marrow
53	6	3.1	38	22	AAW20290	Peptide #6724 enco
54	6	3.1	38	22	AAW34490	Peptide #8527 enco
55	6	3.1	38	23	ABG44284	Human peptide enco
56	6	3.1	39	22	ABG18021	Novel human diagno
57	6	3.1	42	22	ABW31406	Peptide #4057 enco
58	6	3.1	42	22	ABW36620	Peptide #4126 enco
59	6	3.1	42	22	ABW21956	Protein #3955 enco
60	6	3.1	42	22	AAW57381	Human brain expres
61	6	3.1	42	22	AAW69776	Human bone marrow
62	6	3.1	42	22	AAW17600	Peptide #4034 enco
63	6	3.1	42	22	AAW30121	Peptide #4158 enco
64	6	3.1	42	22	AAW05261	Peptide #3943 enco
65	6	3.1	42	23	ABG39406	Human peptide enco
66	6	3.1	48	22	ABG03951	Novel human diagno
67	6	3.1	49	22	ABW29257	Peptide #1908 enco
68	6	3.1	49	22	ABW34422	Peptide #1928 enco
69	6	3.1	49	22	ABW19834	Protein #1833 enco
70	6	3.1	49	22	AAW55212	Human brain expres
71	6	3.1	49	22	AAW67607	Human bone marrow
72	6	3.1	49	22	AAW15414	Peptide #1848 enco
73	6	3.1	49	22	AAW27903	Peptide #1940 enco
74	6	3.1	49	22	AAW03176	Peptide #1858 enco
75	6	3.1	49	23	ABG37197	Human peptide enco
76	6	3.1	51	22	AAU45186	Propionibacterium
77	6	3.1	53	21	AAW53800	Amino acids 145-19
78	6	3.1	54	23	ABP07204	Human ORFX protein
79	6	3.1	56	22	AAW90475	Human immune/haema
80	6	3.1	64	22	AAW55056	Propionibacterium
81	6	3.1	66	21	AAW91503	Human secreted pro
82	6	3.1	67	22	ABG15915	Novel human diagno
83	6	3.1	69	21	AAW15927	Arabidopsis thalia

```
QY 130 LysSerArgSerLys---SerGlnSerSerLysSerGln-----ThrArgAsnAsn 145
Db ::::||||| ||| ||||||| ::::||||| ||| |||||||
311 CGAACCCTCGAANAACGTCCCAACGACAAACAAACACGACCCGACCCGAAACTAC 370
QY 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
Db ::::||||| ||| ||||||| ::::||||| ||| |||||||
371 AACATCAAAACGCGCTAAACACTAAACGGCGACACCAACCAACGCAACGCGTAACCGA 430
QY 166 ArgGlnGluArgHisTle-----LysGluGluGlnAspLys-GluMetTh 180
Db ||| :::: ||| ||| :::: ||| ||| :::: ||| |||
431 CGACTTAACCTCGGAATTACAAAACACCGAAGAACTCCGCGAATAAACTCAAC 490
QY 180 rSerAlaLysGlnHisLeuLeuPheValArgLysAsn 192
Db ||| ::::||||| ||| ||| :::: ||| |||
491 CTCGAACCGACACATCTCTCTCCCTCCCGCGCCTAAC 527
```

Search completed: February 25, 2003, 10:59:34  
Job time : 1267 secs



CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2736 BP; 867 A; 521 C; 608 G; 740 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 0.422 Length: 2736  
Score: 99.00 Matches: 24  
Percent Similarity: 58.57% Conservative: 17  
Best Local Similarity: 34.29% Mismatches: 21  
Query Match: 9.81% Indels: 8  
DB: 24 Gaps: 2  
  
US-09-868-352-23 (1-192) x ABR84078 (1-2736)  
QY 122 AspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSer----- 137  
DB 801 GATAAGAAAGAGAAAGAGGCGCATTCAGATCAAGATCAAGTTCAGGAGGAGGAGG 860  
QY 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPhe 155  
DB 861 ACTCCCTCATCTCTAGACACAGCGGTCAGAGAGGAGATCGAGCGGTCACATTCT 920  
QY 156 ThrSerLysLysArgLysAspThrLys-----ArgArgGlnGluArgHisIleLysGlu 173  
DB 921 AAGTCTAGGAGTCGGCGACGATCCAAAGCCCAAGCGGAGAGATCTCATTCACAGAA 980  
QY 174 GluGlnAspLysGluMetThrSerAlaLys 183  
DB 981 AGAGGTAGAGGTCAAGGAGCATCAAAA 1010  
  
RESULT 38  
AAC78163  
ID AAC78163 standard; cDNA; 3484 BP.  
XX  
AC AAC78163;  
XX  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human cancer associated gene sequence SEQ ID NO:557.  
XX  
KW Human; cancer associated gene; cancer antigen; detection; cancer;  
KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;  
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;  
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;  
KW dermatological; neuroprotective; thrombolytic; coagulant; neutropic;  
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
KW allergic reaction; graft versus host disease; organ rejection;  
KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
KW neurological disease; drug screening; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200055350-A1.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000WO-US05882.  
XX  
XX 12-MAR-1999; 99US-0124270.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM;  
XX  
XX WPI; 2000-587533/55.  
XX  
XX P-PSDB; AAB43954.  
XX  
XX Novel isolated nucleic acids comprising sequences encoding peptides

PT useful for treating or diagnosing e.g. cancer -  
XX  
PS Claim 1; Page 1083-1084; 2352pp; English.  
XX  
XX AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC in AAB43398 to AAB44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vulnery; immunomodulator;  
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;  
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;  
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
CC neutropic; vasotropic; antipsoriatic and antiangiogenic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 3484 BP; 1111 A; 607 C; 695 G; 1068 T; 3 other;  
  
Alignment Scores:  
Pred. No.: 0.58 Length: 3484  
Score: 99.00 Matches: 24  
Percent Similarity: 58.57% Conservative: 17  
Best Local Similarity: 34.29% Mismatches: 21  
Query Match: 9.81% Indels: 8  
DB: 21 Gaps: 2  
  
US-09-868-352-23 (1-192) x AAC78163 (1-3484)  
QY 122 AspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSer----- 137  
DB 1101 GATAAGAAAGAGAAAGAGGCGCATTCAGATCAAGATCAAGTTCAGGAGGAGGAGG 1160  
QY 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPhe 155  
DB 1161 ACTCCCTCATCTCTAGACACAGCGGTCAGAGAGGAGATCGAGCGGTCACATTCT 1220  
QY 156 ThrSerLysLysArgLysAspThrLys-----ArgArgGlnGluArgHisIleLysGlu 173  
DB 1221 AAGTCTAGGAGTCGGCGACGATCCAAAGCCCAAGCGGAGAGATCTCATTCACAGAA 1280  
QY 174 GluGlnAspLysGluMetThrSerAlaLys 183  
DB 1281 AGAGGTAGAGGTCAAGGAGCATCAAAA 1310  
  
RESULT 39  
ABQ50644/c  
ID ABQ50644 standard; DNA; 528 BP.  
XX  
AC ABQ50644;  
XX  
XX 12-JUL-2002 (first entry)  
XX  
XX  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 37235.  
DE  
DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200218632-A2.  
XX  
XX





OS Homo sapiens.  
PN W0200058473-A2.  
XX 05-OCT-2000.  
XX 31-MAR-2000; 2000WO-US08621.  
XX 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
PA Shimkets RA, Leach M;  
XX WPI: 2000-602362/57.  
XX P-PSDB; AAB41766.  
DR Novel nucleic acids and peptides derived from open reading frame X,  
XX useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX Claim 5; Page 2281-2282; 5507pp; English.  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antihydroid; and antanaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS-viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX Sequence 1411 BP; 521 A; 255 C; 329 G; 305 T; 1 other;  
SQ

Alignment Scores:  
Pred. No.: 0.176 Length: 1411  
Score: 99.00 Matches: 29  
Percent Similarity: 46.00% Conservative: 17  
Best Local Similarity: 29.00% Mismatches: 22  
Query Match: 9.81% Indels: 32  
DB: 21 Gaps: 4

US-09-868-352-23 (1-192) x AAC75975 (1-1411)  
QY 115 ProArgAspIleLysPheAspGluArgProHis----- 126  
DB 310 CCTGTCACATTCTTAATAGATAGAACGAGCGATCGATCAAGCAGTAGCTTCTTAT 369  
QY 127 -----LysArgGlySerArgSerLysSerGlnSer----- 137  
DB 370 GGCTCCAGAGGAACCAAGTCAAGTCTCAAGGGGTCGAGGGAATCCTATAGATT 429  
QY 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 153  
DB 430 CAGAGGTCTAGGTCACAAAGACAGAACAGAGGTCCAGGTCAAGACCTCTCCGTCT 489  
QY 154 HisPheThrSerLysLysArg-----LysAspThr 163

Db 490 CATAGTCGTAGCAGTGAAGGTCACAGAACGCGTAGTCGGTCTCGGGATAGA 549  
QY 164 LysArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThrSerAlaLys 183  
Db 550 GAACGACGTAAGGGCAGAGAT---AAAGAGAAAAGAGAAAGAGAGGATAAAGGGAAG 606  
RESULT 36  
ABK83839  
ID ABK83839 standard; cDNA; 2076 BP.  
XX AC ABK83839;  
XX DT 14-AUG-2002 (first entry)  
XX DE Human cDNA differentially expressed in granulocytic cells #410.  
XX KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX OS Homo sapiens.  
XX WO200228999-A2.  
XX PD 11-APR-2002.  
XX PF 03-OCT-2001; 2001WO-US30821.  
XX PR 03-OCT-2000; 2000US-237189P.  
XX PA (GENE-) GENE LOGIC INC.  
XX PI Beazer-Barclay Y, Weissman SM, Yamada S, Vockley J;  
XX WPI: 2002-435328/46.  
XX PT Detecting granulocyte activation by detecting differential expression  
XX of genes associated with granulocyte activation, which serves as  
XX diagnostic markers that is useful for monitoring disease states and  
XX drug toxicity -  
XX PS Claim 1; SEQ ID NO 410; 114pp; English.  
XX CC The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile

KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; periodontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200228999-A2.  
 XX  
 XX PD 11-APR-2002.  
 XX  
 XX PF 03-OCT-2001; 2001WO-US30821.  
 XX  
 XX PR 03-OCT-2000; 2000US-237189P.  
 XX  
 XX PA (GENE-) GENE LOGIC INC.  
 XX  
 XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
 PI WPI; 2002-435328/46.  
 DR  
 XX  
 XX Detecting granulocyte activation by detecting differential expression  
 PT of genes associated with granulocyte activation, which serves as  
 PT diagnostic markers that is useful for monitoring disease states and  
 PT drug toxicity -  
 XX  
 XX Claim 1; SEQ ID No 859; 114pp; English.  
 XX  
 CC The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing  
 CC the expression level to an expression level in an unactivated  
 CC GC, where differential expression of Gs is indicative of GCA.  
 CC Also included are modulating (M2) GA by contacting GC with an agent  
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
 CC for an agent capable of modulating GCA or an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease using the  
 CC gene expression profile; (3) detecting (M4) an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease, by detecting the  
 CC level of expression in a sample of the tissue of gene(s) from Gs, where  
 CC the level of expression of the gene is indicative of inflammation;  
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
 CC an allergic response in a subject, exposure of a subject to a pathogen  
 CC or sterile inflammatory disease, by contacting a tissue having  
 CC inflammation with an agent that modulates the expression of gene(s)  
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
 CC modulating Gs; M3 is useful for screening an agent capable of modulating  
 CC GCA preferably in an inflammation in a tissue; M4 is useful for  
 CC detecting an inflammation (especially chronic) in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
 CC periodontal disease; also bacterial infection, viral infection,  
 CC parasitic infection, protozoal infection, fungal infection and M5 is  
 CC useful for treating one of the above conditions. The present  
 CC sequence represents a gene differentially expressed in granulocytes.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1454 BP; 432 A; 255 C; 365 G; 402 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.161 Length: 1454  
 Score: 99.50 Matches: 43  
 Percent Similarity: 41.23% Conservative: 44  
 Best Local Similarity: 20.38% Mismatches: 73  
 Query Match: 9.86% Indels: 51

DB: 24 Gaps: 8  
 US-09-868-352-23 (1-192) x ABK84288 (1-1454)  
 Qy 20 PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsn 39  
 Db 210 TTTGTGGAATTTGAGGATCCAAAGGATCGAGATGATGCTGTGTATGAGCTTGTGGA 269  
 Qy 40 GluLysSerAlaPheAspValThrVal----- 48  
 Db 270 GAACTCTGTAGTGAAGGGTTACTATTGAACATGCTAGGGCTCGGTACAGAGTGGAAGA 329  
 Qy 49 -----PheGlyGlnArgPheSerGluIleLeuLysTyrAsp----- 61  
 Db 330 GGTAGAGGACGATACTCTGACCGTTTGTAGTAGTCGCACACCTCGAAATGATAGACGAAAT 389  
 Qy 62 -----PheIleValGlyAsp----- 66  
 Db 390 GCTCCACCTGTAGAACAGAAAAATCGTCTTATAGTTGAGAAATTTATCTCAAGATCAGC 449  
 Qy 67 TrpGlyAsnGluGlnLeuArgGlyPheTyrLysAspAlaSerThrIleArg--- 85  
 Db 450 TGGCAGGAT-----CTCAAGATTTCATGAGACAAGCTGGGGAAGTAAACGTTT 497  
 Qy 86 LysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCys 105  
 Db 498 CGGATGTCACACCGACCTAAATTAATGAAGGGTGGTGTGAGTTGCCTCTTATGGTGAC 557  
 Qy 106 AlaTyrPheValLeuGlu-----AsnProAsnProArgAspIleLys--- 119  
 Db 558 TTAAAGAAATGCTATTGAAATAAATCTTGGAAAGGAAATAAATGGAGAAAAATAAAATTA 617  
 Qy 120 PheAspAspGluArgProHisLysArgLysSerArgSerLysSerGlnSer----- 137  
 Db 618 ATTGAAGGCAGCAAGAACGACAGTAGGTCAAGAAAGCAGGTCTCGATCCCGACCAAGAAGT 677  
 Qy 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsn 152  
 Db 678 TCCTCTAGGTCTCGTAGCCGATCCGTTCCGCTAGTCGCAATCTTACAGCCGGTCAAGA 737  
 Qy 153 AlaHisPheThrSerLysLysArgLysAspThrLysArgGlnGluArgHisIleLys 172  
 Db 738 AGCAGGACGAGGACCGGAGCGGAGCAAGTCCCGTTCTGTAGTAGTCTCCCGTGCCT 797  
 Qy 173 GluGluGlnAspLysGluMetThrSerAlaLys 183  
 Db 798 GAGAAGAGCCAGAAACGTTGTTCTTCAAGTAGA 830  
 RESULT 35  
 AAC75975  
 ID AAC75975 standard; CDNA; 1411 BP.  
 XX AAC75975;  
 AC AAC75975;  
 XX  
 DT 08-FEB-2001 (first entry)  
 DE Human ORF1530 polynucleotide sequence SEQ ID NO:3059.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnervary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX

CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations of single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX

SQ Sequence 1236 BP; 157 A; 153 C; 421 G; 505 T; 0 other;

#### Alignment Scores:

Pred. No.: 0.114 Length: 1236  
Score: 100.00 Matches: 22  
Percent Similarity: 61.25% Conservative: 27  
Best Local Similarity: 27.50% Mismatches: 25  
Query Match: 9.91% Indels: 6  
DB: 24 Gaps: 3

US-09-868-352-23 (1-192) x ABQ29237 (1-1236)

Qy 110 LeuGluAsnProAsnProArgAspIleLys-----PheAspAspGluArgProHisLys 127  
Db 1235 ATCGAAATCGAAATCGAAAAATCGTAAACGCGGATACGATAATCGCTACAAAAATAAT 1176  
Qy 128 ArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSer 147  
Db 1175 CGACGCGAACCAATAATCGTGAATAATAATCGACGCGAATCT---AATCGACGCGCTA 1119  
Qy 148 GlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArgArgGln 167  
Db 1118 AAAAAATATCGAAAT-----ACGTTCGCGCGCGAAAAATAATCGACGAAAAACC 1068  
Qy 168 GluArgHisIleLysGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuLeu 187  
Db 1067 GAACGACGCGCTAAAAATAATAAAAAAATCCAAACATCGCCCAAAAAAATAATCTC 1008

#### RESULT 33

ID ABQ29237 standard; DNA; 1236 BP.  
AC ABQ29237;  
XX  
XX 12-JUL-2002 (first entry)  
DT  
DE  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 15828.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
XX 07-MAR-2002.  
PD  
XX  
XX 01-SEP-2001; 2001WO-EPI0074.  
PF  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
PR  
XX  
XX 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX

PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -  
XX

PS Claim 12: 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations of single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX

SQ Sequence 1236 BP; 505 A; 421 C; 153 G; 157 T; 0 other;

#### Alignment Scores:

Pred. No.: 0.114 Length: 1236  
Score: 100.00 Matches: 22  
Percent Similarity: 61.25% Conservative: 27  
Best Local Similarity: 27.50% Mismatches: 25  
Query Match: 9.91% Indels: 6  
DB: 24 Gaps: 3

US-09-868-352-23 (1-192) x ABQ29237 (1-1236)

Qy 110 LeuGluAsnProAsnProArgAspIleLys-----PheAspAspGluArgProHisLys 127  
Db 2 ATCGAAATCGAAATCGAAAAATCGTAAACGCGGATACGATAATCGCTACAAAAATAAT 61  
Qy 128 ArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSer 147  
Db 62 CGACGCGAACCAATAATCGTGAATAATAATCGACGCGAATCT---AATCGACGCGCTA 118  
Qy 148 GlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArgArgGln 167  
Db 119 AAAAAATATCGAAAT-----ACGTTCGCGCGCGAAAAAATAATCGACGAAAAACC 169  
Qy 168 GluArgHisIleLysGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuLeu 187  
Db 170 GAACGACGCGCTAAAAATAATAAAAAAATCCAAACATCGCCCAAAAAAATAATCTC 229

#### RESULT 34

ABK84288  
ID ABK84288 standard; cDNA; 1454 BP.  
XX  
XX ABK84288;  
AC  
XX  
XX 14-AUG-2002 (first entry)  
DT  
XX  
XX Human cDNA differentially expressed in granulocytic cells #859.  
DE  
XX  
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW

KW glomerulonephritis; multiple sclerosis; diabetes mellitus; osteoporosis;  
 KW cancer; cirrhosis; hepatitis; psoriasis; transgenic animal; antitumor;  
 KW tranquilliser; drug screening; pancreatitis; renal tubular acidosis;  
 KW systemic lupus erythematosus; colitis; ss.

OS Homo sapiens.

FH Key Location/Qualifiers  
 FT 64..1938  
 CDS /\*tag= a  
 FT /product= "Human RMEP-20 protein"

XX WO200183524-A2.

XX 08-NOV-2001.

XX 27-APR-2001; 2001WO-US13862.

XX 28-APR-2000; 2000US-200184P.

PR 04-MAY-2000; 2000US-201875P.

PR 04-MAY-2000; 2000US-202090P.

PR 06-JUN-2000; 2000US-210232P.

PR 25-JUL-2000; 2000US-220553P.

XX (INCY-) INCYTE GENOMICS INC.

XX Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J, Hillman JJ;  
 PI Baughn MR, Yao MG, Burford N, Batra S, Policky JJ;

XX WPI; 2002-034502/04.

DR P-PSDB; AAE15257.

XX New human RNA metabolism protein for diagnosing or treating nervous  
 PT system disorders, autoimmune/inflammatory disorders, cell proliferative  
 PT disorders and developmental disorders

XX Claim 11; Page 180-181; 196pp; English.

XX The invention relates to human RNA metabolism proteins (RMEP) and their  
 CC corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,  
 CC treating and preventing nervous system disorders (epilepsy, dementia,  
 CC stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);  
 CC prion diseases; fatal familial insomnia, nutritional and metabolic  
 CC diseases of the nervous system; inherited, metabolic, endocrine and  
 CC toxic myopathy; mental disorders (mood, anxiety, schizophrenic disorders)  
 CC amnesia and Tourette's disorder; autoimmune/inflammatory disorders (AIDS-  
 CC acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,  
 CC atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,  
 CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
 CC arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,  
 CC ulcerative colitis, and infections); cell proliferative disorders (cancer  
 CC arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental  
 CC disorders (renal tubular acidosis). RMEP DNA is useful in drug screening  
 CC techniques, gene therapy and for creating transgenic animals. The present  
 CC sequence is human RMEP-20 cDNA.

XX Sequence 2220 BP; 789 A; 436 C; 555 G; 440 T; 0 other;

Alignment Scores:

Pred. No.:	0.19	Length:	2220
Score:	101.00	Matches:	46
Percent Similarity:	43.83%	Conservative:	25
Best Local Similarity:	28.40%	Mismatches:	71
Query Match:	10.01%	Indels:	21
DB:	24	Gaps:	6

US-09-868-352-23 (1-192) x AAD24393 (1-2220)

Qy 32 IleGluPhe-----GlnLeuValIleAsnGluLysSerAlaPheAspValThrValPhe 49

Db 742 GTGGATTTCAGACCAAAATTCGTACCAAGGGCCCTGCTTTAATGGAGTAGTTT 801

Qy 50 GlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsn 69

Db	802	GGAGACAGGCCACTCAAAATAAATCACTCCACAATGCAATAGTAAACCC-----	852
Qy	70	GluGlnLeuArgLeuArgGlyPheTyrIlyAspAlaSerThrIleArgLysAsnSerArg	89
Db	853	CCTGAGATGCACACCTCAGGCTGCAGCTAAGGAGTTAGAAGAGTAATGAAGCGA-----	906
Qy	90	IleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheVal	109
Db	907	-----GTACGAGAAGCTCAGTCATTATTCTCAGCAGCTATTGAA	945
Qy	110	LeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgProHisLysArgArg	129
Db	946	CCAGAGTCTGGAAGAGCAATGAAGAAAGCGGCTCGATCTCGTTCCCATCTCGCTCA	1005
Qy	130	LysSerArgSerLysSerGlnSerSerLysGlnThrArgAsnAsnArgSerGlnSer	149
Db	1006	AAATCCAGGTCT-----AGTCAAAATCCCATTCTAGAAGAAAGATCAATCA	1056
Qy	150	-----AsnAlaAsnAlaHisPheThrSerLysLysArg---LysAspThrLysArgArg	166
Db	1057	AAACACAGGAGTAGATCCCATTAATAGATCAGCTTCAAGACAGAAAGACAGAGTAGATCT	1116
Qy	167	GlnGluArgHisIleLysGluGlnAspLysGluMetThrSerAlaLysGlnHisLeu	186
Db	1117	AAGAGCCACATAAAACGCTCTAAATCAAGGGAGAG-ACGGAAGTCAAGGAGTCGTTTC	1175
Qy	187	LeuPhe 188	
Db	1176	GCATTTC 1181	
RESULT 32			
ABQ29236/C			
ID	ABQ29236	standard; DNA; 1236 BP.	
XX	AC	ABQ29236;	
XX	AC		
DT	12-JUL-2002	(first entry)	
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 15827.		
XX	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;		
KW	drug; side effect; cancer; central nervous system; cardiovascular;		
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;		
KW	SNP; cell differentiation; ds.		
XX	Homo sapiens.		
XX	WO200218632-A2.		
PD	07-MAR-2002.		
XX	01-SEP-2001; 2001WO-EPI0074.		
XX	01-SEP-2000; 2000DE-1043826.		
PR	05-SEP-2000; 2000DE-1044543.		
XX	(EPIG-) EPIGENOMICS AG.		
XX	Olek A, Piepenbrock C, Berlin K, Guetig D;		
PI	WPI; 2002-371829/40.		
DR	Determining the degree of cytosine methylation in genomic DNA, useful		
PT	for diagnosis and prognosis, comprises selective hybridization of		
PT	amplicons from chemically treated DNA		
XX	Claim 12; 56pp + Sequence Listing; 56pp; German.		
XX	This invention describes a novel method for determining the degree of		
CC	methylation of a particular cytosine in a motif 5'-CpG-3', present in a		
CC	genomic sample of DNA. The sample is treated chemically to convert		
CC	cytosine (C) but not methylated C, to uracil, then part of the genomic		

US-09-868-352-23 (1-192) x AB035135 (1-886)

QY 112 AsnProAsnProArgAspIleLysPheAspGluArgProHisLysArgArgLysSer 131  
 DB 448 AACACGAATCCACGCTACTCGACGTATCGTCTCGTCCCAACGCTAGGACTAATCA 507

QY 132 ArgSerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 151  
 DB 508 ACGCGCTCTCCAAATAACGCGCGGACACACGCGACAAACGCGTCTCGCTTAACCC 567

QY 152 AsnAlaHisPheThrSerLysArgLysAspThrLysArgGlnGluArgHisIle 171  
 DB 568 AATAATGACCTCTCTACCAAAAGAAACGACCACTCGTCGAACACGTACAAAT--- 624

QY 172 LysGluGlnGluAspLysGluMetThrSerAlaLysGlnHis 185  
 DB 625 ---AACCAATACGATAAAATAACAACAACCTCAAAAAAATC 663

RESULT 30  
 ID AAA70121  
 AC AAA70121 standard; DNA: 1527 BP.  
 XX  
 XX AAA70121;  
 DT 07-NOV-2000 (first entry)

XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:254.  
 DE  
 XX Plasmodium falciparum; chromosome 2: human malaria parasite; vaccine;  
 KW antimalarial; malaria; protozoa; infection; insecticide; ds.  
 XX Plasmodium falciparum.  
 OS  
 XX WO200025728-A2.  
 PN  
 XX 11-MAY-2000.  
 PD  
 XX 05-NOV-1999; 99WO-US26796.  
 PF  
 XX 05-NOV-1998; 98US-0107131.  
 PR  
 XX (HOFF/) HOFFMAN S.  
 PA (CARU/) CARUCCI D.  
 PA (GARD/) GARDNER M.  
 PA (VENT/) VENTER J C.  
 XX  
 XX Hoffman S, Carucci D, Gardner M, Venter JC;  
 PI  
 XX WPI; 2000-365347/31.  
 DR  
 XX  
 XX Proteins encoded by chromosome 2 of the human malarial parasite,  
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
 PT diagnosis of P.falciparum infection.  
 PS  
 XX Disclosure; Page 472; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
 CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
 CC vaccines against P. falciparum infection comprising (I) or (II).  
 CC (I) and (II) are useful for the development of vaccines against  
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
 CC antibody raised to immunogens comprising the sequences of (I), are  
 CC useful in the detection of infection with P. falciparum. Furthermore,  
 CC (I) (especially when they are rifins or secreted or membrane proteins)  
 CC can aid the identification of drugs to treat or prevent P. falciparum  
 CC infection, or they can be used to identify drug resistance in  
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
 CC subsequent identification of proteins encoded by it will help to expand  
 CC our understanding of parasite biology, a process hampered by the  
 CC complexity of the parasitic lifecycle, and provide new targets for  
 CC vaccine and drug development. Parasite resistance to drugs and mosquito  
 CC resistance to insecticides have led to a resurgence of malaria in many

CC parts of the world, and there is a pressing need for vaccines and new  
 CC drugs. AAAT0078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
 CC and protein sequences given in the present invention, but which are not  
 CC specifically mentioned within the specification.

SQ Sequence 1527 BP; 848 A; 94 C; 210 G; 375 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.116 Length: 1527  
 Score: 101.00 Matches: 41  
 Percent Similarity: 40.31% Conservative: 38  
 Best Local Similarity: 20.92% Mismatches: 61  
 Query Match: 10.01% Indels: 56  
 DB: 21 Gaps: 9

US-09-868-352-23 (1-192) x AAA70121 (1-1527)

QY 5 ValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPheIleHisPheGlu 24  
 DB 790 ATAACAAAAGAA---TGTAATGTGTAATAATATGAT-----GAT 825

QY 25 AsnIleVal-----LysSerAspAspIleGlu--- 33  
 DB 826 AATATTATAGATCACAACAAACACAGGAGAAAGAAAGAAAGAAAGTATCGAAAT 885

QY 34 -----PheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50  
 DB 886 ATCAATATAAGCCATATATATGAAAGAAAGAAACATCTCATGACATTTCT----- 936

QY 51 GlnArgPheSerGluIleLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70  
 DB 937 -----AATGTATTAGAAGAAATAAAGAG 960

QY 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIle 90  
 DB 961 GAAGAAAATATACAAATTTACAAAAGAGATGTTATACAAATTTGTAATATGTAAGGTA 1020

QY 91 SerArgLeuGluAspTyr-----IleLysGluTyrCysAsnPheGlyCysAlaTyrPhe 108  
 DB 1021 ---AAACTTGAGAAATATCATCATGAAAGAAATTTGAAATAAT----- 1059

QY 109 ValLeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgProHisLysArg 128  
 DB 1060 ---GTTCAAAATTAATAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1116

QY 129 ArgLysSerArgSerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGln 148  
 DB 1117 AATAAAAAAGGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1176

QY 149 SerAsnAlaAsnAlaHisPheThrSerLysArgLysAspThrLysArgArgGlnGlu 168  
 DB 1177 -----GAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1209

QY 169 ArgHisIleLysGluGlnGluAspLysGluMetThrSerAlaLysGln 184  
 DB 1210 AAAAATAAATAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 1257

RESULT 31  
 ID AAD24393  
 XX AAD24393 standard; cDNA: 2220 BP.  
 AC AAD24393;  
 DT 07-MAR-2002 (first entry)  
 DE Human RNA metabolism protein-20 (RMEP-20) cDNA.  
 KW Human; RNA metabolism protein-20; RMEP-20; gout; nervous system disorder;  
 KW autoimmune; inflammatory; cell proliferative; developmental; thyroiditis;  
 KW gene therapy; epilepsy; dementia; stroke; Alzheimer's disease; amnesia;  
 KW Parkinson's disease; prion disease; insomnia; endocrine disorder; AIDS;  
 KW Acquired Immune Deficiency Syndrome; mental disorder; allergy; anaemia;  
 KW asthma; atherosclerosis; Crohn's disease; rheumatoid arthritis; vaccine;

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; KW gastrointestinal; respiratory system; single nucleotide polymorphism; KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
XX  
PD 07-MAR-2002.  
XX  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
PR  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
XX (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
XX WPI; 2002-371829/40.  
XX  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of PT amplicons from chemically treated DNA.  
XX  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
XX This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. AB013410-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

XX SQ Sequence 886 BP; 114 A; 123 C; 309 G; 340 T; 0 other;

Alignment Scores:  
Pred. No.: 0.0565 Length: 886  
Score: 101.00 Matches: 24  
Percent Similarity: 52.70% Conservative: 15  
Best Local Similarity: 32.43% Mismatches: 33  
Query Match: 10.01% Indels: 2  
DB: 24 Gaps: 1

US-09-868-352-23 (1-192) x ABQ35134 (1-886)

Qy 112 AsnProAsnProArgAspIleLysPheAspGluArgProHisLysArgArgLysSer 131  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 439 AACAGAAATCCACGGTACTGACGTATCGCTATCTCTGCTCCCAACGTAACGACTAAATCA 380  
  
Qy 132 ArgSerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 151  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 379 ACCGCTCTCCAAATACCGCGCACACACGCGTCTCTCGCTAAACCC 320  
  
Qy 152 AsnAlaHisPheThrSerLysLysArgLysAspThrLysArgGlnGluArgHisIle 171  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 319 AATAATCGACCTCTTAACCAAAAGCAACGCACTCGTCCGAACACGTACAAAT--- 263

Qy 172 LysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 262 --AACCAATACGATAAAATACACAACTCAAAAAAACAAC 224  
  
RESULT 29  
ABQ35135  
ID ABQ35135 standard; DNA; 886 BP.  
XX  
AC ABQ35135;  
XX  
XX 12-JUL-2002 (first entry)  
DE  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21726.  
XX  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; KW gastrointestinal; respiratory system; single nucleotide polymorphism; KW SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200218632-A2.  
XX  
XX 07-MAR-2002.  
XX  
XX  
XX 01-SEP-2001; 2001WO-EP10074.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
PR  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
XX (EPiG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
XX WPI; 2002-371829/40.  
XX  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of PT amplicons from chemically treated DNA.  
XX  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. AB013410-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

XX SQ Sequence 886 BP; 340 A; 309 C; 123 G; 114 T; 0 other;

Alignment Scores:  
Pred. No.: 0.0565 Length: 886  
Score: 101.00 Matches: 24  
Percent Similarity: 52.70% Conservative: 15  
Best Local Similarity: 32.43% Mismatches: 33  
Query Match: 10.01% Indels: 2  
DB: 24 Gaps: 1

[illegible]





KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnerary; antiulcer; ss.  
 XX Homo sapiens.  
 XX WO200157188-A2.  
 XX 09-AUG-2001.  
 XX 05-FEB-2001; 2001WO-US03800.  
 XX 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 XX (HYSE-) HYSEQ INC.  
 XX PA  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX WI: 2001-457740/49.  
 DR P-PSDB; ABB11354.  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 XX Claim 1; Page 469-470; 1963pp; English.  
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a cDNA encoding a  
 CC novel human polypeptide of the invention.  
 XX  
 SQ Sequence 1909 BP; 687 A; 340 C; 461 G; 415 T; 6 other;

## Alignment Scores:

Pred. No.: 0.105 Length: 1909  
 Score: 102.50 Matches: 33  
 Percent Similarity: 48.96% Conservative: 14  
 Best Local Similarity: 34.38% Mismatches: 36

Query Match: 10.16% Indels: 14  
 DB: 22 Gaps: 4  
 US-09-868-352-23 (1-192) x ABA08598 (1-1909)  
 QY 103 PheGlyCysAla-----TyrPheValLeuGluAsnProAsnPro 115  
 DB 933 TTTAATTGTTCTGAGCTTACACTGCAAAAGTGATTTTCTCCAGAGTCTCGAAGAGC 992  
 QY 116 ArgAspIleLysPheAspGluArgProHisLysArgGlnLysSerArgSerLysSer 135  
 DB 993 AATGAAGAAAAGGCGGTGCGATCTGTTCCCATATCTCGTCAAAATCCAGGTCT----- 1046  
 QY 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSer-----AsnAlaAsnAla 153  
 DB 1047 ---AGCTCAAAATCCCATTTCTAGAGGAAAGAACATCACAATCAAAACACAGGAGTAGATCC 1103  
 QY 154 HisPheThrSerLysLysArg---LysAspThrLysArgGlnGlnArgHisLysLys 172  
 DB 1104 CATAATAGATCAGTTCAGACAGAAAGACAGACGTAGATCTTAAGAGCCCATATAAAAA 1163  
 QY 173 GluGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuLeuPhe 188  
 DB 1164 CGCTCTAAATCAAGGAGAG-ACGGAAGTCAAGGAGTCTGTCGCATTC 1210  
 RESULT 25  
 AAD39557  
 ID AAD39557 standard; DNA; 2695 BP.  
 XX AC AAD39557;  
 XX DT 04-OCT-2002 (first entry)  
 XX DE Human SR-cyp DNA.  
 XX KW Human; antisense: SR-cyp; Clk-associated RS cyclophilin; inflammation;  
 KW hyperproliferative disorder; cancer; prophylaxis; infection; therapy;  
 KW tumour; CARs-cyp; gene; ds.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT CDS 158..2422  
 FT /\*tag= a  
 FT /product= "Human SR-cyp protein"  
 XX PN WO200236809-A2.  
 XX PD 10-MAY-2002.  
 XX PF 30-OCT-2001; 2001WO-US47335.  
 XX PR 03-NOV-2000; 2000US-0706197.  
 XX XX (ISIS-) ISIS PHARM INC.  
 XX PA (COLD-) COLD SPRING HARBOR LAB.  
 XX PI Bennett CF, Spector DL, Wyatt JR;  
 XX DR WPI: 2002-479763/51.  
 XX DR P-PSDB; AAE24596.  
 XX PT Novel antisense compounds targeted to nucleic acids encoding SR-cyp,  
 PT Clk-associated RS cyclophilin for modulating the gene expression and  
 PT treating hyperproliferative disorders such as cancer -  
 XX PS Example 13; Page 95-99; 117pp; English.  
 XX  
 CC The invention relates to antisense compounds targeted to a nucleic  
 CC acid molecule encoding human SR-cyp (Clk-associated RS cyclophilin)  
 CC to inhibit its expression. SR-cyp is also referred to as CARs-cyp.  
 CC Antisense compounds of the invention are used for treating diseases  
 CC or conditions associated with SR-cyp. The diseases treated include





Qy 54 SerGluIleLeuLeuLysTyrAspPheIleValGlyAspTTPGlyAsnGluInLeuArg 73  
Db 86 TCCGACATTTTAAATAACGACTACATCTCGGGGATTGGGATACAGCCAGCTGAGG 145  
Qy 74 LeuArgGlyPheTyrLysAspAlaSer---ThrIleArgLysAsnSerArgIleSerArg 92  
Db 146 CTGAAGGCTTTTGTGAGCATCAAAATCAAAAGGCCACATTTGTATACGAAGATCAGCAGC 205  
Qy 93 LeuGluAspTyrIleLysTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111  
Db 206 CTGTATGAATATATTACGAGTACTGCAATTTGCGATGCGCTACTTTGTTGTTTAA 262

RESULT 20  
AAX84569  
ID AAX84569 standard; cdna to mRNA; 6145 BP.  
XX  
AC AAX84569;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE Biorhythm marker gene.  
XX  
KW Biorhythm marker gene; circadian rhythm; ss.  
XX  
OS Rattus sp.  
XX  
PN JP11169185-A.  
XX  
PD 29-JUN-1999.  
XX  
PF 12-DEC-1997; 97JP-0362890.  
XX  
PR 12-DEC-1997; 97JP-0362890.  
XX  
PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.  
PA (SAKA ) OTSUKA PHARM CO LTD.  
XX  
XX WPI; 1999-422623/36.  
DR P-PSDB; AAY22206.  
XX  
XX Structure and function of a new biorhythm marker gene - and protein encoded by it  
XX  
XX Claim 7; Page 14-16; 23pp; Japanese.  
XX  
CC This sequence encodes the biorhythm marker protein of the invention. The  
CC protein is also referred to as a circadian rhythm protein.  
XX  
SQ Sequence 6145 BP; 1849 A; 1110 C; 1364 G; 1822 T; 0 other;

Alignment Scores:  
Pred. No.: 0.0277 Length: 6145  
Score: 113.50 Matches: 55  
Percent Similarity: 37.16% Conservative: 42  
Best Local Similarity: 21.07% Mismatches: 73  
Query Match: 11.25% Indels: 91  
DB: Gaps: 11

US-09-868-352-23 (1-192) x AAX84569 (1-6145)

Qy 8 GluMetLeuAsnTyrAsnLysTyrProGlyProGlnPheIleHisPheGluAsnIleVal 27  
Db 2788 AAATGTTAATGTAAAGCAGGTTCCTTCCCTCAAGTTTAAATTTTCAGACAAGAA 2847  
Qy 28 LysSerAspIleGlu----- 33  
Db 2848 AGGCACTCTAATTTAGAAAACCAAGATATTCAGTTTCTAGACTAAAGGAAAACTTACT 2907  
Qy 34 -----PheGlnLeuValIleAsnGluLysSerAlaPhe-----AspValThr 47  
Db 2908 CAAATTTTGTGTGCTAAATACATTTAACTCCACTCTGAAGTCTCTGATTTATCT 2967  
Qy 48 ValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPhe----- 62

Db 2968 TTCATT-----AGATTTCCTAAATAATACTTGAAATTTAGTTTCGGTTCTAACAAAAATA 3021  
Qy 63 -----IleValGlyAsp 66  
Db 3022 AATCAGGAAAGTCTTACGCCTCACTGCTCTAGGAATTGCAGACATGTTTAACTCTGCGT 3081  
Qy 67 TrpGlyAsnGluInLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys 86  
Db 3082 TATGGCAGCAACAGAAATA---CTAGAACAGCTGTTTATAATCATCTCGTTTATATGATAC 3138  
Qy 87 AsnSerArgIleSerArgLeuGluAsp-----TyrIleLys 98  
Db 3139 TGCTGCAGGGTGGCTGCAGCTCAACGAGTTTATGCAATGACTTTCTTGGATGTTTCTGAAG 3196  
Qy 99 GluTyrCysAsnPheGly----- 104  
Db 3199 GATAATTGCACAGGAGGAGGATGTACAGAGAGTAGGCCCTTGCACTATATGTGTATAT 3258  
Qy 105 -----CysAlaTyr----- 107  
Db 3259 TCCACTTGTGCCTATTATTAACTGGGATCTTTAATTGTTCTGAGCTTACGCTGCAAAAGT 3318  
Qy 108 -----PheValLeuGluAsnProArgAsnProArgAspIleLysPheAspAspGluArgPro 125  
Db 3319 GGTGTTTCTCCAGAGTCTGGAAAGAGCAATCAAGGAAGGGCGTCGATCTCGTTCC 3378  
Qy 126 HisLysArgArgLysSerArgSerLysSerGlnSerLysSerGlnThrArgAsnAsn 145  
Db 3379 CACACTCGCTCAAAAGTCCAGTCT-----AGCTCAAAAGTCCCATTTCTAGAAGCAAG 3429  
Qy 146 ArgSerGlnSer-----AsnAlaAsnAlaHisPheThrSerLysLysArg---LysAsp 162  
Db 3430 AGATCGCAGTCAAAAGCAGGAGTAGATCCCTAACAGGTCCAGTTCAGAGCAAGAGAT 3489  
Qy 163 ThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSerAla 182  
Db 3490 AGAGTAGATCCAAAGAGCCCAACATAAGAAAGCGTCTTAAGTCTAGGAGAGCGGAAGTCA 3549  
Qy 183 Lys 183  
Db 3550 AGG 3552

RESULT 21  
AA158597  
ID AA158597 standard; cdna; 1826 BP.  
XX  
AC AA158597;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 800.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW ankyrotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX  
XX 25-APR-2000; 2000US-0552317.  
XX  
XX 09-JUL-2000; 2000US-0598042.  
XX  
XX 19-JUL-2000; 2000US-0620312.  
XX  
XX 03-AUG-2000; 2000US-0653450.

PI Berka R, Clausen IG;  
XX WPI; 2002-416684/44.  
XX Monitoring differential expression of several genes in first Bacillus  
PT cell relative to expression of same genes in one or more second  
PT Bacillus cells, by using substrate containing Bacillus genomic  
PT sequenced tag array -  
XX  
XX Claim 11: SEQ ID NO 7936; 200pp; English.  
XX  
XX The invention describes a method of monitoring differential expression of  
CC genes in a first Bacillus cell relative to expression of the genes in  
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
CC isolated from Bacillus cells to a substrate containing array of Bacillus  
CC genomic sequenced tags (GST), examining the array, and determining  
CC relative gene expression by an observed hybridisation reporter signal of  
CC a spot in the array. The method is useful for measuring the expression of  
CC genes in a first Bacillus cell relative to expression of the same genes  
CC in one or more second Bacillus cells. The method is useful for monitoring  
CC global expression of several genes from a Bacillus cell, discovering new  
CC genes, identifying possible functions of unknown open reading frames and  
CC monitoring gene copy number variation and stability. Monitoring changes  
CC in expression of genes may be used to provide a representation of the way  
CC in which Bacillus cells adapt to changes in culture conditions,  
CC environmental stress or other physiological provocation. Extensive  
CC follow-up characterisation is unnecessary, when one spot on an array  
CC equals one gene or one open reading frame, since sequence information is  
CC available. This sequence represents a genomic sequence tag (GST) used in  
CC the method of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 285 BP; 93 A; 51 C; 66 G; 74 T; 1 other;  
XX  
Alignment Scores:  
Pred. No.: 1.46e-17 Length: 285  
Score: 232.50 Matches: 46  
Percent Similarity: 69.32% Conservative: 15  
Best Local Similarity: 52.27% Mismatches: 22  
Query Match: 23.04% Indels: 5  
DB: 24 Gaps: 2  
US-09-868-352-23 (1-192) x ABK80645 (1-285)  
QY 26 IleVallySerAspApilGluPheGlnLeuValIleAsnGluLysSerAlaPheAsp 45  
Db 1 ATGGTGACGATTCAGCGATTCATGATATGATGCTCTCAAAACGAGCGCATGTTGGAAT 60  
QY 46 ValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGly 65  
Db 61 GAGGAGGCATTCAAAGCGCGTACAGCGATGATCAACAAATATGATTATCATCGTCGCG 120  
QY 66 AspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp-----AlaSer 82  
Db 121 GATTGGGGCTACACCAACGCTGCGGCTTCTTTCGATGACACAAACAAATCA 180  
QY 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102  
Db 181 ACATTCGATAAA-----AAATAGACACGCTTACCCGACATCTTTACGATACTGCAAT 234  
QY 103 PheGlyCysAlaTyrPheValLeu 110  
Db 235 TTTGGTTCGCTTATTTGCTTG 258  
RESULT 19  
ABK75854  
ID ABK75854 standard; DNA; 273 BP.  
XX  
AC ABK75854;  
XX

DT 13-AUG-2002 (first entry)  
XX  
DE Bacillus licheniformis genomic sequence tag (GST) #3145.  
XX  
KW Differential gene expression; genomic sequenced tag; GST;  
KW altered culture condition; environmental stress;  
KW physiological provocation; ds.  
XX  
OS Bacillus licheniformis.  
XX  
PN WO200229113-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-US31437.  
XX  
PR 06-OCT-2000; 2000US-0680598.  
XX  
PT 27-MAR-2001; 2001US-279526P.  
XX  
PA (NOVO ) NOVOZYMES BIOTECH INC.  
PA (NOVO ) NOVOZYMES AS.  
XX  
PI Berka R, Clausen IG;  
XX  
XX WPI; 2002-416684/44.  
DR  
XX Monitoring differential expression of several genes in first Bacillus  
PT cell relative to expression of same genes in one or more second  
PT Bacillus cells, by using substrate containing Bacillus genomic  
PT sequenced tag array -  
XX  
XX Claim 4; SEQ ID NO 3145; 200pp; English.  
XX  
XX The invention describes a method of monitoring differential expression of  
CC genes in a first Bacillus cell relative to expression of the genes in  
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
CC isolated from Bacillus cells to a substrate containing array of Bacillus  
CC genomic sequenced tags (GST), examining the array, and determining  
CC relative gene expression by an observed hybridisation reporter signal of  
CC a spot in the array. The method is useful for measuring the expression of  
CC genes in a first Bacillus cell relative to expression of the same genes  
CC in one or more second Bacillus cells. The method is useful for monitoring  
CC global expression of several genes from a Bacillus cell, discovering new  
CC genes, identifying possible functions of unknown open reading frames and  
CC monitoring gene copy number variation and stability. Monitoring changes  
CC in expression of genes may be used to provide a representation of the way  
CC in which Bacillus cells adapt to changes in culture conditions,  
CC environmental stress or other physiological provocation. Extensive  
CC follow-up characterisation is unnecessary, when one spot on an array  
CC equals one gene or one open reading frame, since sequence information is  
CC available. This sequence represents a genomic sequence tag (GST) used in  
CC the method of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 273 BP; 89 A; 48 C; 61 G; 75 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 1.89e-16 Length: 273  
Score: 222.50 Matches: 42  
Percent Similarity: 74.68% Conservative: 17  
Best Local Similarity: 53.16% Mismatches: 19  
Query Match: 22.05% Indels: 1  
DB: 24 Gaps: 1  
US-09-868-352-23 (1-192) x ABK75854 (1-273)  
QY 34 PheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPhe 53  
Db 26 TTTGAGTCGTCRAAGATGTCAAAACGGTTTTAACGAAGACGCTTAAAGCCAGATAC 85

CC anti-Listeria vaccines.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1234 BP; 402 A; 195 C; 288 G; 349 T; 0 other;  
Alignment Scores:  
Pred. No.: 1-62e-17 Length: 1234  
Score: 239.50 Matches: 43  
Percent Similarity: 77.38% Conservative: 22  
Best Local Similarity: 51.19% Mismatches: 14  
Query Match: 23.74% Indels: 5  
DB: 24 Gaps: 2  
US-09-868-352-23 (1-192) x ABQ68718 (1-1234)  
Qy 31 AspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50  
Db 541 GATTTAAACTACGAGATTATCACCAATTATCGCGACGCTTTTGACGAGGAGAAAGTTAAAT 600  
Qy 51 GlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70  
Db 601 GAACGGTTTAGCGATATCTTGGCGGATATGATTATATAGTAGCGACTGGGGTTACGAT 660  
Qy 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys-----Asn 87  
Db 661 CAACCTCCGCTCAAGGCTTTTGAAGACGATAAC-----CGCAAAGCCGCATACGAC 714  
Qy 88 SerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyr 107  
Db 715 AACAAATAGCACGCTTAAAGAAATATATTACGAATACTGCAATTTTGGTTGTGCGTAT 774  
Qy 108 PheValLeuGlu 111  
Db 775 TTTGTTATTAAA 786  
RESULT 17  
ABQ70820  
ID ABQ70820 standard; DNA; 2690 BP.  
XX AC ABQ70820;  
XX  
DT 29-AUG-2002 (first entry)  
XX  
DE Listeria monocytogenes 4b contig DNA sequence #762.  
XX  
KW Antibacterial; Listeria; food contamination; mutational analysis;  
KW infection; ds.  
XX  
OS Listeria monocytogenes 4b.  
XX  
PN WO200228891-A2.  
XX  
PD 11-APR-2002.  
XX  
PE 04-OCT-2001; 2001WO-FR03061.  
XX  
PR 04-OCT-2000; 2000FR-0012697.  
XX  
PA (INSP ) INST PASTEUR.  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Kunst F, Glaser P;  
XX  
DR WPI; 2002-332479/37.  
XX  
PT New genomic sequences from Listeria species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators  
XX  
PS Claim 14; SEQ ID 3633; 180pp; French.  
XX

CC The present invention relates to nucleic acid sequences  
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes  
CC and primers for identification and/or detection of Listeria (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of Listeria (potential therapeutic agents), also for  
CC treating infections by Listeria, and are useful as immunogens in  
CC anti-Listeria vaccines.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2690 BP; 852 A; 456 C; 629 G; 753 T; 0 other;  
Alignment Scores:  
Pred. No.: 4.54e-17 Length: 2690  
Score: 239.50 Matches: 43  
Percent Similarity: 77.38% Conservative: 22  
Best Local Similarity: 51.19% Mismatches: 14  
Query Match: 23.74% Indels: 5  
DB: 24 Gaps: 2  
US-09-868-352-23 (1-192) x ABQ70820 (1-2690)  
Qy 31 AspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50  
Db 541 GATTTAAACTACGAGATTATCACCAATTATCGCGACGCTTTTGACGAGGAGAAAGTTAAAT 600  
Qy 51 GlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70  
Db 601 GAACGGTTTAGCGATATCTTGGCGGATATGATTATATAGTAGCGACTGGGGTTACGAT 660  
Qy 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys-----Asn 87  
Db 661 CAACCTCCGCTCAAGGCTTTTGAAGACGATAAC-----CGCAAAGCCGCATACGAC 714  
Qy 88 SerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyr 107  
Db 715 AACAAATAGCACGCTTAAAGAAATATATTACGAATACTGCAATTTTGGTTGTGCGTAT 774  
Qy 108 PheValLeuGlu 111  
Db 775 TTTGTTATTAAA 786  
RESULT 18  
ABK80645  
ID ABK80645 standard; DNA; 285 BP.  
XX AC ABK80645;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Bacillus clausii genomic sequence tag (GST) #3488.  
XX  
KW Differential gene expression; genomic sequenced tag; GST;  
KW altered culture condition; environmental stress;  
KW physiological provocation; ds.  
XX  
OS Bacillus clausii.  
XX  
PN WO200229113-A2.  
XX  
PD 11-APR-2002.  
XX  
PR 05-OCT-2001; 2001WO-US31437.  
XX  
PR 06-OCT-2000; 2000US-0680598.  
PR 27-MAR-2001; 2001US-279526P.  
XX  
PA (NOVO ) NOVOZYMES BIOTECH INC.  
PA (NOVO ) NOVOZYMES AS.  
XX

Db 241 TGTCCTTATTGTTACTCAGACGATTGTCACCAATGAATTTATTGAGAAATAGATGAT 300  
QY 123 Glu 123  
DB: :::  
Db 301 AAA 303  
RESULT 15  
ID ABA03041/c  
ID ABA03041 standard; DNA; 2944528 BP.  
XX ABA03041;  
XX AC  
DT 05-FEB-2002 (first entry)  
XX  
XX Listeria monocytogenes EGD-e genome sequence.  
XX  
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease; ds.  
KW  
XX Listeria monocytogenes.  
OS  
XX  
XX WO200177335-A2.  
XX  
XX 18-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-FR01118.  
XX  
XX 11-APR-2000; 2000FR-0004629.  
XX  
XX (INSP ) INST PASTEUR.  
XX  
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;  
XX  
XX WPI; 2002-010914/01.  
XX  
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides -  
PT  
XX  
PS Claim 1; SEQ ID No 1; 192pp; French.  
XX  
XX The present sequence is the genome sequence of Listeria monocytogenes  
CC Egd-e. This sequence and fragments of this sequence are useful for  
CC selecting probes and primers for detecting genes in L. monocytogenes and  
CC related organisms, and to study genetic polymorphisms and other genomes.  
CC Proteins (ABB47297-ABB50149) expressed from the present sequence are  
CC useful for raising specific antibodies, identification of L.  
CC monocytogenes and related organisms, and for biosynthesis of  
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and  
CC proteins encoded by it are also useful for selecting compounds that  
CC regulate gene expression and cell replication and modulate L.  
CC monocytogenes-related diseases. In addition, this sequence and proteins  
CC encoded by it are useful in pharmaceutical and vaccines compositions for  
CC the treatment or prevention of infections by L. monocytogenes and related  
CC organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Alignment Scores:  
Pred. No.: 3-58e-13 Length: 2944528  
Score: 240.50 Matches: 53  
Percent Similarity: 67.23% Conservative: 27  
Best Local Similarity: 44.54% Mismatches: 28

Query Match: 23.84% Indels: 11  
DB: 24 Gaps: 4  
US-09-868-352-23 (1-192) x ABA03041 (1-2944528)  
QY 2 ArgLysGluVal---ThrProGluMetLeu--AsnTyrAsnLysTyrProGlyProGlnP 20  
DB 2475946 CGAAAGAAGTATATACACCGGACTACTCCGAGATATTTTAAATGGAACCTCAAAAA 2475887  
QY 20 heIleHisPhe-----GluAsnIleValLysSerAspIleGluPheGlnLeuV 37  
DB 2475886 ATGTACGACGAGGAGCAAGATGACGATTACGATTCAAGATTTAACATACGAGATTA 2475827  
QY 37 aIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleL 57  
DB 2475826 TCACCAATATTATCGCATGCTTTTGACGAGGAGAACTTAAATGAACGGTTAGCGATATTC 2475767  
QY 57 euLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyP 77  
DB 2475766 TTGGACGATATGATTATATAGTGGCGACTGGGTTACGATCAACTCGGCTGAAAGCT 2475707  
QY 77 heTyrLysAspAlaSerThrIleArgLys-----AsnSerArgIleSerArgLeuG 94  
DB 2475706 TTTTGAAGACGATAAC-----CGCAAGCCGCATACGACACAAATATTACACGTTAA 2475653  
QY 94 luAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111  
DB 2475652 AAGAAATATATTATGAACTACTGCAATTTTGGTGTGCGTATTTTGTATTATAA 2475600  
RESULT 16  
ABO68718  
ID ABO68718 standard; DNA; 1234 BP.  
XX AC ABO68718;  
XX  
DT 29-AUG-2002 (first entry)  
XX  
DE Listeria monocytogenes 4b contig DNA sequence #1484.  
XX  
KW Antibacterial; Listeria; food contamination; mutational analysis;  
KW infection; ds.  
XX  
OS Listeria monocytogenes 4b.  
XX  
PN WO200228891-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 04-OCT-2001; 2001WO-FR03061.  
XX  
PR 04-OCT-2000; 2000FR-0012697.  
XX  
PA (INSP ) INST PASTEUR.  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Kunst F, Glaser P;  
XX WPI; 2002-332479/37.  
XX  
XX New genomic sequences from Listeria species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators -  
XX  
PS Claim 14; SEQ ID 1531; 180pp; French.  
XX  
XX The present invention relates to nucleic acid sequences  
CC (ABO67188-ABO71212) from Listeria sp. The sequences are useful as probes  
CC and primers for identification and/or detection of Listeria (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of Listeria (potential therapeutic agents), also for  
CC treating infections by Listeria, and are useful as immunogens in





KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
XX Staphylococcus aureus.  
OS  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 1141..1200  
FT /\*tag= a  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
FT misc\_feature 2941..3000  
FT /\*tag= b  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
FT misc\_feature 4741..4800  
FT /\*tag= c  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
FT misc\_feature 6541..6600  
FT /\*tag= d  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
XX  
PN EF786519-A2.  
XX  
XX 30-JUL-1997.  
XX  
XX 07-JAN-1997; 97EP-0100117.  
XX  
XX 05-JAN-1996; 96US-0009861.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
XX Rosen CA;  
XX  
XX WPI; 1997-374922/35.  
XX  
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
XX stored on computer readable medium and used in the production of  
XX anti-S.aureus vaccines  
XX  
XX Claim 1; Page 958-962; 3271pp; English.  
XX  
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
XX of the invention. The DNA sequences are recorded on a computer readable  
XX medium, preferably selected from a floppy or hard disk, random access  
XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
XX the S.aureus DNA sequences allows putative functions to be assigned so  
XX that protein-encoding or regulatory regions of commercial, therapeutic or  
XX industrial importance can be obtained. Specifically, sequences which are  
XX likely to encode antigens have been identified and these polypeptides can  
XX be used in a vaccine composition against S.aureus infection. The  
XX polypeptides can also be used in a kit for the immunodetection of  
XX S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
XX including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
XX skin and surgical wound infections, scalded skin syndrome, toxic shock  
XX syndrome, etc. Organisms transformed with the DNA sequences can be used  
XX for recombinant production of the polypeptides. The new DNA sequences  
XX (and their fragments) are useful as primers or probes for isolating  
XX homologues of any of the S.aureus DNA sequences contained on the  
XX computer readable medium.

SQ Sequence 7769 BP; 2697 A; 1099 C; 1425 G; 2307 T; 241 other;  
Alignment Scores:  
Pred. No.: 1.35e-17 Length: 7769  
Score: 249.50 Matches: 50  
Percent Similarity: 62.28% Conservative: 21  
Best Local Similarity: 43.88% Mismatches: 30  
Query Match: 24.73% Indels: 13  
DB: 18 Gaps: 2  
US-09-868-352-23 (1-192) x AAV74525 (1-7769)  
QY 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAla 43  
Db 6606 AAGAATTTGATAAAAGATAGATCAACATTTACTTTGAATTAATAGAAATATCGCGAATGT 6665  
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIle 63  
Db 6666 TTTAATGAAGAACAATTTTGTCTAGGTATTCAGATATTTTAGATAAATATGATTACATA 6725  
QY 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSer--- 82  
Db 6726 GTTGTGTACTATGCTTACGATCAATTTACGATTAAGAGTTTTCACAAAGATTCTTAATAAA 6785  
QY 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102  
Db 6786 AAGCAGAGATGAGTAACGTTTTTCAAAATATTCAAGATTACATATTTGAATATTGTAAC 6845  
QY 103 PheGlyCysAlaTyrPheValLeu----- 110  
Db 6846 TTTGGTTGTCCTTACCTTTGATTAAAGACATTTGCTCTAAACAGAGGTTAAAGATTAAATC 6905  
QY 111 GluAsnProAsnProArgAspIleLysPheAspGluArg 124  
Db 6906 GAAGAAAGTTTCATCGTCTGATGTGATAGATGACGACAATAAA 6947  
RESULT 13  
ABQ69245/c  
ID ABQ69245 standard; DNA; 3011208 BP.  
XX  
XX AC ABQ69245;  
XX  
XX 29-AUG-2002 (first entry)  
XX  
XX DE Listeria innocua DNA sequence #684.  
XX  
XX KW Antibacterial; Listeria; food contamination; mutational analysis;  
XX infection; ds.  
XX  
XX OS Listeria innocua.  
XX  
XX PN WO200228891-A2.  
XX  
XX PD 11-APR-2002.  
XX  
XX PF 04-OCT-2001; 2001WO-FR03061.  
XX  
XX PR 04-OCT-2000; 2000FR-0012697.  
XX  
XX PA (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX PI Kunst F, Glaser P;  
XX WPI; 2002-332479/37.  
XX  
XX DR  
XX PT New genomic sequences from Listeria species, useful for detection,  
XX treatment and prevention of infection, also related polypeptides,  
XX antibodies and modulators -  
XX  
XX PS Claim 5; SEQ ID 2058; 180pp; French.  
XX  
XX CC The present invention relates to nucleic acid sequences

CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX  
 SQ Sequence 3054 BP; 955 A; 532 C; 440 G; 1127 T; 0 other;

Alignment Scores:  
 Pred. No.: 8.17e-19 Length: 3054  
 Score: 255.50 Matches: 48  
 Percent Similarity: 68.22% Conservative: 25  
 Best Local Similarity: 44.86% Mismatches: 31  
 Query Match: 25.32% Indels: 3  
 DB: 22 Gaps: 2

US-09-868-352-23 (1-192) x AAH54843 (1-3054)

Qy 20 PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsn 39  
 Db 2328 TTTTGTGAAGTGAATAATGATTAAAGTCGACCAACAATATTTTGAATTGATAGAGAA 2269  
 Qy 40 GluLysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLys 59  
 Db 2268 TATAGACAGATTTTGTATGACGAATAATTTTCACCTAGGTATTCGGATATATAGACAAA 2209  
 Qy 60 TyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLys 79  
 Db 2208 TATGATTATGTCGTAGTACTATGTTACGATCAATTACGCTTAAAGGATTTTATATA 2149  
 Qy 80 AsplaserThrIleArgLys--AsnSerArgIleSerArgLeuGluAspTyrIleLys 98  
 Db 2148 GATAGTAAATAAAAGCGCAAAATAAGTAAACGATTTTCAAGTATACAGATATATATACTA 2089  
 Qy 99 GluTyrCysAsnGlyGlyAlaTyrPheValLeuGluAsnProAsnProArgAspIle 118  
 Db 2088 GAATATTGTAATTTGGTGTCTTATTTGTACTACAGCATTTGTCACCAATGAATTT 2029  
 Qy 119 -----LysPheAspGlu 123  
 Db 2028 ATTTGAAGAAATAGATGATAA 2008

RESULT 11  
 ABN92594  
 ID ABN92594 standard; DNA; 399 BP.  
 XX  
 AC ABN92594;  
 XX  
 XX  
 DT 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2057.  
 XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy; gene; ds.  
 XX Staphylococcus epidermidis.  
 OS  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX

PR 14-AUG-1997; 97US-055779p.  
 PR 08-NOV-1997; 97US-064964p.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 DR WPI; 2002-381255/41.  
 DR P-PSDB; ABP40049.  
 XX  
 PT Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 XX  
 PS Disclosure; SEQ ID 2057; 267pp; English.  
 XX  
 CC ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly *S. epidermidis* infections. The sequences can be used to  
 CC screen for compounds able to interfere with the *S. epidermidis* life  
 CC cycle or inhibit *S. epidermidis* infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 SQ Sequence 399 BP; 162 A; 41 C; 70 G; 126 T; 0 other;

Alignment Scores:  
 Pred. No.: 7.24e-20 Length: 399  
 Score: 254.50 Matches: 47  
 Percent Similarity: 68.93% Conservative: 24  
 Best Local Similarity: 45.63% Mismatches: 29  
 Query Match: 25.22% Indels: 3  
 DB: 24 Gaps: 2

US-09-868-352-23 (1-192) x ABN92594 (1-399)

Qy 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAla 43  
 Db 4 GAAATATGATGATAAGTCGACCAACAATATTTTGAATTGATAGAGAAATGT 63  
 Qy 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIle 63  
 Db 64 TTTGATCAGGAATATTTTCAGCTAGGTATTCGGATATATAGACAAATATGATTATGTC 123  
 Qy 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr 83  
 Db 124 GTAGGTGACTATGCTTACGATCAATTACGCTAAAGGATTTTATAAGATAGTAATAA 183  
 Qy 84 IleArgLys--AsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102  
 Db 184 AAGCGAGAAATAAGTAAACGATTTTCAAGTATATACAGATTTATATAGTAATTTGTAAT 243  
 Qy 103 PheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle-----LysPhe 120  
 Db 244 TTTGGTGTCTTATTTTGTAGTCAGACGATTTGTCACCAATGAATTTTATGAGAAATA 303  
 Qy 121 AspAspGlu 123  
 Db 304 GATGATAAA 312

RESULT 12  
 AA74525  
 ID AA74525 standard; DNA; 7769 BP.  
 XX  
 AC AA74525;  
 XX  
 DT 16-MAR-1999 (first entry)  
 XX Staphylococcus aureus contig SEQ ID #214.  
 DE  
 XX



CC from the present invention can be used in diagnosis kits and assays, and  
CC pharmaceutical compositions and vaccines for *S. pneumoniae*.  
XX  
SQ Sequence 5558 BP; 1628 A; 1239 C; 1067 G; 1624 T; 0 other;

(INRG ) INRA INST NAT RECH AGRONOMIQUE.

PA  
XX  
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
XX  
XX WPI; 2002-043418/06.  
XX  
XX  
XX New nucleotide sequence useful in the identification of *Lactococcus*  
PT  
PT lactis and related species -  
XX  
XX  
XX Claim 1; SEQ ID 1; 2504pp; French.  
XX  
XX The present invention is related to a *Lactococcus lactis* nucleotide  
CC  
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
CC  
CC nucleic acid sequence is useful in the detection and/or amplification of  
CC  
CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or  
CC  
CC related species. The proteins of the invention are useful for the  
CC  
CC biosynthesis or biodegradation of a composition of interest. The  
CC  
CC invention helps research in lactic bacteria, particularly useful in the  
CC  
CC production of yogurt and cheese.  
CC  
CC Note: The sequence data for this patent is based on equivalent patent  
CC  
CC W0200177334 (published 18-OCT-2001) which is available in electronic  
CC  
CC format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
XX  
XX  
XX Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence encodes an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

31-OCT-1996: 96US-0029960.

XX	(HUMA-) HUMAN GENOME SCI INC.
PA	
XX	
XX	Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI	Kunsch CA, Rosen CA;
PI	
XX	WPT; 1998-272225/24.
DR	
XX	
XX	Computer-readable medium with recorded Streptococcus pneumoniae
PT	polynucleotide sequences - useful in diagnostic kits and assays, and
PT	pharmaceutical compositions and vaccines for Streptococcus
PT	pneumoniae
XX	
XX	Claim 1; Page 770-773; 1409pp; English.
PS	

compact reduce medium with received Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
XX  
PS Claim 1; Page 770-773; 1409pp: English.

Claim 1; Page 770-773; 1409pp; English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from *Streptococcus pneumoniae*. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the *S. pneumoniae* genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridise to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the *S. pneumoniae* genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the *S. pneumoniae* genome of commercial importance, or expression modulating fragments of the *S. pneumoniae* genome. Products

```

Db 562 CATAAAGAAAGTCA---AAACGTAATCAGACTAGTCAACTTAATACCAAAATTAAGTCAT 618
QY 186 LeuLeuPheValArgLysAsn 192
Db 619 TTTATCATCAGAAAGAAAGAT 639

RESULT 5
AAS55679
ID AAS55679 standard; DNA; 531 BP.
AC AAS55679;
XX
XX
XX 13-FEB-2002 (first entry)
XX Streptococcus pneumoniae DNA for cellular proliferation protein #250.
XX Antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX Streptococcus pneumoniae.
XX WO200170955-A2.
XX 27-SEP-2001.
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR P-PSDB; AAU37820.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 9316; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 531 BP; 190 A; 92 C; 118 G; 131 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 3.42e-42 Length: 531

```

```

Score: 452.50 Matches: 94
Percent Similarity: 66.49% Conservative: 33
Best Local Similarity: 49.21% Mismatches: 48
Query Match: 44.85% Indels: 16
DB: 23 Gaps: 4

US-09-868-352-23 (1-192) x AAS55679 (1-531)
QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
Db 1 ATGCGAAAAGAAATTCACCTGAATTATACAACAACTTTCCTGGTCCGAGATTC 60
QY 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
Db 61 CATTTACACGGGACAGGTGGAACGGAAGGATAGCTTTTTCCTTGGTGAATAATATC 120
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
Db 121 AAGGATGCCTTTGATGTGACGACTTTTAAATCAGCGCTTTTTCAGAAAGTATTAACCAAGTTT 180
QY 61 AspPheIleValGlyAspTrpClyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
Db 181 GATTATATCGTGGGGACTGGAGCAACGACGCTTCGCCCTACGAGGTTTTTACAGGAT 240
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
Db 241 GACCGAACAGAGAAACAACTTCAAAAAATCAGTCCTTTACAGACTACTCTTTAGAGTAT 300
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
Db 301 TGTAGTTATGGTTGTCCTATTCTTCTAGAAAAATGAAGCCCTTAAGCGAGCATCATTT 360
QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSer 140
Db 361 GAC-----AAGAAATCGGTAAGACGGAAGAACACCTTCTAGAAAAAGGAAGAAA 414
QY 141 GlnThrArgAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
Db 415 CCGACTCAAACTAAACGA---AAGTCGATTCGAGAT----- 447
QY 161 LysAspThrLysArgArgGlnGlnArgHisIleLysGluGlnAspLysGluMetThr 180
Db 448 AAGAAAAATAGACGTCGTCAGAAAGACCGACGATTTCTCAGAAAGAGGACAAAGGAA 501
QY 181 SerAlaLysGlnHis-LeuLeuPheValArg 190
Db 502 -----CAACGTCATTTTGTTCATTCGTCAGA 526

RESULT 6
AAS55832
ID AAS55832 standard; DNA; 531 BP.
XX
XX AC AAS55832;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE Streptococcus pneumoniae DNA for cellular proliferation protein #403.
XX
XX Antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX Streptococcus pneumoniae.
XX WO200170955-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 21-MAR-2001; 2001WO-US09180.
XX
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.

```

Db 1 ATCGGAAAGAGTACACAGAGATGCTTAACATATAAAGTATCTCTGCCCCACAGTTC 60  
 QY 21 ILeHisPheGluAsnIleValIysSerAspIleGluPheGlnLeuValIleAsnGlu 40  
 Db 61 ATTCACCTTCAAAATATCGTTAAAGTGATGATATTGAATTTCAACTTGTATTAAATGAA 120  
 QY 41 LysSerAlaPheAspValThrValPheGlyClnArgPheSerGluIleLeuLeuLysTyr 60  
 Db 121 AAATACAGCTTTTGTATGATGCTGCTTTGGACAACGTTTCTCAGATTTTATTAAATAT 180  
 QY 61 AspPheIleValGlyAspTTPGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80  
 Db 181 GATTTATCGTTGGCGATTCGGGTAGACACAGTTGAGGCTAGAGCGCTTTTCAAGAT 240  
 QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100  
 Db 241 GCTAGTACAATTAGAAAAATAGCGGATTTACGTTTACAGATTATATTAAAGAGTAT 300  
 QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120  
 Db 301 TGTAACTTTTGGTTGTGCTTATTGTTGGAGAAATCCAAATCCTAGAGATATTAAATTT 360  
 QY 121 AspAspGluArgProHisLysArgLysSerArgLysSerGlnSerLysSer 140  
 Db 361 GATGATCAAGACCTCATAGCGTCTGTAAGTCAAGATCCAAATCACAAATCATCAAGTCA 420  
 QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLysLysArg 160  
 Db 421 CAAACTAGAAATAATCGTCCAGTCAAAATGCCAATGCTCATTTTACAAGTAAAGCGT 480  
 QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180  
 Db 481 AAAGACAAAGCCGCTCAAGAACGTCATATTAAAGAGCAAGCAAGATAAGGAATGACC 540  
 QY 181 SerAlaLysGlnHis-LeuLeuPheValArgLys 191  
 Db 541 TCTGCAAGCAGCATTTTGTATTCTGTAAGAAA 574  
 RESULT 4  
 ID ABN66118 standard; DNA: 642 BP.  
 XX AC ABN66118;  
 XX AC  
 XX AC  
 DT 01-JUL-2002 (first entry)  
 DE Streptococcus polynucleotide SEQ ID NO 149.  
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 XX OS Streptococcus pyogenes.  
 XX XX  
 XX XX  
 PN W0200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 XX 29-OCT-2001; 2001WO-GB04789.  
 XX  
 PR 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 PA  
 PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 XX WPI: 2002-352536/38.  
 DR P-PSDB: ABP25487.  
 XX

PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX Claim 7; Page 3168; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX  
 XX Sequence 642 BP; 251 A; 106 C; 114 G; 171 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 4, 01e-54 Length: 642  
 Score: 558.50 Matches: 107  
 Percent Similarity: 71.98% Conservative: 42  
 Best Local Similarity: 51.69% Mismatches: 41  
 Query Match: 55.35% Indels: 17  
 DB: 24 Gaps: 4  
 US-09-868-352-23 (1-192) x ABN66118 (1-642)  
 QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20  
 Db 25 ATGAAAAAGAAAAATTTGCCGAGAAATGTACACTATAATAATTTCCAGGTCCTCAAAATTT 84  
 QY 21 ILeHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40  
 Db 85 ATTCAATTTGAAGAGCAAGTTAAGGCTGAAGCATTTGTTTCTCTAGAAAGATGTT 144  
 QY 41 LysSerAlaPheAspValThrValPheGlyClnArgPheSerGluIleLeuLeuLysTyr 60  
 Db 145 AAGAAGCGCTTTTGATACGACTAGCTTTGGTCAAGCTTATACAGAAGTCTCTCTAAAGTAT 204  
 QY 61 AspPheIleValGlyAspTTPGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80  
 Db 205 GACTATATTGTCGGATTCGGGAAATGAACAGCTTCGCTCTTAAAGGGTTTATAAGGAT 264  
 QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100  
 Db 265 AGTCACCATATCAAGAAAAACGAATCGCATCTCAGTTTGAAGATTATATAAGAATTT 324  
 QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120  
 Db 325 TGCAATTTTGGTTGTGCTTATTGTTTGAAGAAATCTTCATCCACAAGATATTAAATTT 384  
 QY 121 AspAspGluArgProHisLysArgLysSer----ArgSerLysSerGln----- 136  
 Db 385 GAAGAGGAGCGCCCAACCAAGCAAGAAATACCTTAATCAAAATCAAAATCGTCGCAAG 444  
 QY 137 -----SerSerLysSerGlnThrArgAsnAsn 145  
 Db 445 CCAAACTATTCAATCAGACGCCGCCAACACCTTAAGAGCAAAATCGAAG---CGGGCATCA 501  
 QY 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165  
 Db 502 AAAGAAAAGCAACCTGAAACCAACCAAGCGCTTTACCAGTCAAAACGTCGAAGTACTACTAAG 561  
 QY 166 ArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSerAlaLysGlnHis 185

```
XX Claim 8; Page 4196-4488; 4525pp; English.
PS The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Alignment Scores:
Pred. No.: 3.65e-99 Length: 2155561
Score: 996.00 Matches: 192
Percent Similarity: 99.48% Conservative: 0
Best Local Similarity: 99.48% Mismatches: 0
Query Match: 98.71% Indels: 1
DB: 24 Gaps: 0

US-09-868-352-23 (1-192) x ABN71527 (1-2155561)
Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
Db 442722 ATGCCAAAAGAGTACACACAGAGATGCTTAACATAATAAGTATCCGCGCCACAGTTC 442781
Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
Db 442782 ATTCACCTTGAATAATCGTTAAAGATGATGATATGAATTCACATGTTTATTAAAGAA 442841
Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyr 60
Db 442842 AAATCAGCTTTTGATGTGACTGTCTTTGGACAACGTTTTCTGAGATTTTATAAAATAT 442901
Qy 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
Db 442902 GATTTTATCGTTGGCGATTGGGGTAACGAGCAGTTGAGCGCTTAAGAGGCTTTTACAAAGAT 442961
Qy 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
Db 442962 GCTAGTACAATTAGAAAAATAGCGGATTTTCACGTTTAGAAGATTTATTTAAAGAGTAT 443021
Qy 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
Db 443022 TGTAACTTTGGTGTGCTTATTTGTGTGGAGAATCCAAATCCCTAGAGATATTAAATTT 443081
Qy 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSer 140
Db 443082 GATGATGAAGACCTTCATAGCGTCTGTAAGTCAGATCCCAATCACAATCATCAAGTCA 443141
Qy 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
Db 443142 CAACTAGAAATAATCGTTCACAGTCAAAATGCCAATGCTCATTTTACAAGTAAAGCGT 443201
Qy 161 LysAspThrLysArgArgGlnGluArgHisLysLysGluGlnAspLysGluMetThr 180
Db 443202 AAAACACAAAACGCCGTCAAGACGTCATATTAAAGAAGAGCAAGATGAAGAAATGACC 443261
Qy 181 SerAlaLysGlnHis-LeuLeuPheValArgLysAsn 192
Db 443262 TCTGCAAGACGACATTTTGTATTTCGTGAAGAAAAAAT 443298
```

```
RESULT 3
ABN66117
ID ABN66117 standard; DNA; 576 BP.
XX
AC ABN66117;
XX
DT 01-JUL-2002 (first entry)
DE Streptococcus polynucleotide SEQ ID NO 147.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI: 2002-352536/38.
DR P-PSDB: ABP25486.
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 7; Page 3168; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 576 BP; 212 A; 86 C; 108 G; 170 T; 0 other;

Alignment Scores:
Pred. No.: 3.38e-103 Length: 576
Score: 990.00 Matches: 191
Percent Similarity: 99.48% Conservative: 0
Best Local Similarity: 99.48% Mismatches: 0
Query Match: 98.12% Indels: 1
DB: 24 Gaps: 0

US-09-868-352-23 (1-192) x ABN66117 (1-576)
Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||
|||||
```



/product= N-acetyl\_muramidase\_homologue

FT XX WO200037646-A2.  
 PN XX 29-JUN-2000.  
 PD XX 22-DEC-1999; 99WO-GB04377.  
 XX XX  
 PR 22-DEC-1998; 98GB-0028345.  
 PR 22-DEC-1998; 98GB-0028349.  
 PR 22-DEC-1998; 98GB-0028350.  
 PR 22-DEC-1998; 98GB-0028352.  
 PR 22-DEC-1998; 98GB-0028353.  
 PR 22-DEC-1998; 98GB-0028354.  
 PR 22-DEC-1998; 98GB-0028355.  
 PR 22-DEC-1998; 98GB-0028356.  
 PR 22-DEC-1998; 98GB-0028357.  
 PR 22-DEC-1998; 98GB-0028359.  
 PR 04-JAN-1999; 99GB-0000082.  
 PR 04-JAN-1999; 99GB-0000083.  
 PR 04-JAN-1999; 99GB-0000084.  
 PR 04-JAN-1999; 99GB-0000085.  
 PR 04-JAN-1999; 99GB-0000086.  
 PR 28-JAN-1999; 99GB-0001916.  
 PR 28-JAN-1999; 99GB-0001922.  
 XX XX  
 PA (MICR-) MICROSCIENCE LTD.  
 XX XX  
 PI Hughes MJG, Santangelo JD, Lane JD, Everest P, Feldman R;  
 PI Moore JC, Wilson RK, Dobson RJ, Dougan G;  
 XX XX  
 DR WPI; 2000-442674/38.  
 XX P-PSDB; AAY96805.  
 XX XX  
 PT New peptides useful for treatment and prevention of conditions  
 PT associated with Streptococcal infection are obtained from group B  
 PT Streptococcus  
 XX XX  
 PS Example 11; Page 46-47; 63pp; English.  
 XX XX  
 CC The S. agalactiae pho3-1 gene encodes a protein which has homology with  
 CC proteins of unknown function in S. pyogenes, S. pneumoniae,  
 CC Bacillus subtilis (yutD) and Enterococcs faecillis. The B. subtilis yutD  
 CC gene is located in a chromosomal region containing genes involved in  
 CC cell wall synthesis.  
 CC Peptides derived from group B Streptococcus (Streptococcus agalactiae)  
 CC proteins encoded by genes pho1-13, pho3-21, pho2-15, pho3-18, pho3-22,  
 CC pho3-3, pho3-17, pho2-2, pho1-5, pho3-1, pho3-23, pho3-50, pho1-14,  
 CC pho2-10, pho3-14, pho3-24 and pho3-29. The peptides are useful for  
 CC screening potential drugs, or for the detection of virulence, and for the  
 CC manufacture of a medicament for use in the treatment or prevention of  
 CC infections such as focal infection (including osteomyelitis, septic  
 CC arthritis, abscesses and endophthalmitis) and urinary tract infections  
 CC caused by group B Streptococcus. The peptides and vaccines comprising the  
 CC peptides are useful in treatment of chronic mastitis, especially in cows  
 CC (i.e. for veterinary purposes).  
 XX XX  
 SQ Sequence 579 BP; 213 A; 85 C; 109 G; 172 T; 0 other;

Alignment Scores:  
 Pred. No.: 2,37e-105 Length: 579  
 Score: 1009.00 Matches: 192  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-868-352-23 (1-192) x AAA51368 (1-579)

QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20  
 Db 1 ATCGGAAAGAGTGACACCAGAGATGCTTAAGTATTAAGTATCTCTGCCCCACAGTTT 60

QY 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40  
 Db 61 ATTCACTTTGAAAAATATCGTTAAAGTGATGATATTGAATTTCAACTTGTATTATGAA 120  
 QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60  
 Db 121 AAATCAGCTTTTGATGTGACTGCTTTTGACACACGTTTTTCTGAGATTTTATTAAAAAT 180  
 QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80  
 Db 181 GATTTTATCGTTGGCGATTGGGGTAACGAGCAGTTCAGGCTAAGAGGCTTTTACAAGAT 240  
 QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100  
 Db 241 GCTAGTACGATTAGAAAAAATAGCCGGATTTACGTTTAGAAGATTATATTAAGAGAT 300  
 QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120  
 Db 301 TGTAACTTTGGTTGCTTATTTTGTGGAGAATCCAAATCCTAGAGATATTAATTT 360  
 QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerLysSer 140  
 Db 361 GATGATGAAAGACCTCATAGCGTCGTAAAGTCAAGATCCAAATCACAATCATCAAGTCA 420  
 QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysArg 160  
 Db 421 CAAACTAGAAATAATCGTTCCCAAGTCAATGCCAATGCTCATTTTACAAGTAAAAAGCGT 480  
 QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180  
 Db 481 AAAGACACAAAACCCGCTCAAGAACGTCATATTAAAGAGAGCAAGATAAGGAATGACC 540  
 QY 181 SerAlaLysGlnHisLeuLeuPheValArgLysAsn 192  
 Db 541 TCTGCAAGCAGCATTTGTTATTTCGTAAGAAAAAAT 576  
 RESULT 2  
 ABN71527  
 ID ABN71527 standard; DNA; 2155561 BP.  
 XX AC  
 XX ABN71527;  
 XX DT  
 XX 02-JUL-2002 (first entry)  
 XX DE Streptococcus polynucleotide SEQ ID NO 10967.  
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 XX OS Streptococcus sp.  
 XX PN WO200234771-A2.  
 XX PD  
 XX 02-MAY-2002.  
 XX PF 29-OCT-2001; 2001WO-GB04789.  
 XX PR 27-OCT-2000; 2000GB-0026333.  
 XX PR 24-NOV-2000; 2000GB-0028727.  
 XX PR 07-MAR-2001; 2001GB-0005640.  
 XX PA (CHIR-) CHIRON SPA.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
 XX PI Tettelin H;  
 XX WPI; 2002-352536/38.  
 XX PT New Streptococcus protein for the treatment or prevention of infection  
 XX or disease caused by Streptococcus bacteria, such as meningitis, and  
 XX PT for detecting a compound that binds to the protein -

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 09:34:32 ; Search time 227 Seconds  
(without alignments)  
1904.774 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 1009

Sequence: 1 MRKEVTPMLNKNYKPGPQF.....EEQDKEMTSKQHLFLVRKN 192

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q/cgn2.1/USPTO.spool/US09868352/runat\_14022003\_153515\_4212/app\_query.fasta\_1.391  
-DB=N\_Geneseq\_101002 -QEXT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868352 -CGN\_1\_187=runat\_14022003\_153515\_4212 -NCPUS=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1009	100.0	579	21	AAA51368 Streptococcus agal
2	996	98.7	2155561	24	ABN71527 Streptococcus poly
3	990	98.1	576	24	ABN66117 Streptococcus poly
4	558.5	55.4	642	24	ABN66118 Streptococcus pneu
5	452.5	44.8	531	23	AA555679 Streptococcus pneu
6	452.5	44.8	531	23	AA555832 Streptococcus pneu
c 7	438	43.4	5558	19	AAV52236 Genomic sequence o
8	331.5	32.9	2365589	24	ABA90521 Enterococcus faeca
9	302.5	30.0	1495	20	AAH13439 S. epidermidis gen
c 10	255.5	25.3	3054	22	AAH54843 Staphylococcus epi
11	254.5	25.2	399	24	ABN92594 Staphylococcus aur
12	249.5	24.7	7769	18	AAV74525 Listeria innocua D
c 13	248	24.6	3011208	24	ABO69245 S. epidermidis ope
14	243.5	24.1	390	22	AAH52964 Listeria monocytog
c 15	240.5	23.8	2944528	24	ABA03041 Listeria monocytog
16	239.5	23.7	1234	24	ABQ68718 Listeria monocytog
17	239.5	23.7	2690	24	ABQ70820 Bacillus clausii g
18	232.5	23.0	285	24	ABK80645 Bacillus lichenifo
19	222.5	22.1	273	24	ABK75854 Biorhythm marker g
20	113.5	11.2	6145	20	AAH84569 Human polynucleoti
21	108.5	10.8	1826	22	AAH58597 DNA encoding novel
22	105	10.4	1808	23	AAH79353 Human polynucleoti
23	102.5	10.2	1583	22	AAI60383 Human Ser/Arg rich
24	102.5	10.2	1909	22	ABA08598 Human SR-cyp DNA.
25	102	10.1	2695	24	AAH39557 Tumour suppressor
26	102	10.1	10369	22	AAH46304 Human immune syste
27	102	10.1	10369	24	ABU32393 Oligonucleotide fo
c 28	101	10.0	886	24	ABQ35134 Oligonucleotide fo
29	101	10.0	886	24	ABQ35135 Plasmodium faicpa
30	101	10.0	1527	21	AAH70121 Human RNA metaboli
31	101	10.0	2220	24	AAH24393 Oligonucleotide fo
c 32	100	9.9	1236	24	ABQ29236 Oligonucleotide fo
33	100	9.9	1236	24	ABQ29237 Human cDNA differe
34	99.5	9.9	1454	24	ABR84288 Human ORFX ORF1530
35	99	9.8	1411	21	AAH75975 Human cDNA differe
36	99	9.8	2076	24	ABK83839 Human cDNA differe
37	99	9.8	2736	24	ABK84078 Human cancer assoc
38	99	9.8	3484	21	AAH78163 Oligonucleotide fo
c 39	98.5	9.8	528	24	ABQ50644 Oligonucleotide fo
40	98.5	9.8	528	24	ABQ50645 Oligonucleotide fo
c 41	98.5	9.8	528	24	ABQ36486 Oligonucleotide fo
42	98.5	9.8	597	24	ABQ36487 Drosophila melanog
43	98.5	9.8	1011	23	ABU24287 Drosophila melanog
c 44	98.5	9.8	3011	23	ABU24286 Human ORFX ORF1661
45	98	9.7	1536	21	AAH76106

ALIGNMENTS

RESULT 1  
AAA51368  
ID AAA51368 standard; DNA; 579 BP.  
XX  
AC AAA51368;  
XX  
DT 26-SEP-2000 (first entry)  
XX  
DE Streptococcus agalactiae pho3-1 gene.  
KW Group B Streptococcus; pho3-1; virulence; vaccine; uropathic;  
KW gene therapy; veterinary; mastitis; immunostimulant; antibacterial; ds.  
XX Streptococcus agalactiae.  
OS  
XX Key Location/Qualifiers  
XX CDS 1..579  
XX FT /\*tag= a

**THIS PAGE BLANK (USPTO)**

```
Best Local Similarity: 22.04%      Mismatches: 66
Query Match: 7.93%                Indels: 58
DB: 4                             Gaps: 7

US-09-868-352-23 (1-192) x US-09-134-001C-2695 (1-789)

Qy 4 GluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPheIleHisPhe 23
   |||:::  ::  |||  |||||  |||  |||
Db 307 GAAATTCACGACGTATTTCCCAATGTTGGAATAAT-----AATTTTATACCTTT 357

Qy 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAla 43
   |||:::  |||||:::  |||:::  |||:::  |||:::  |||:::
Db 358 GAAGTCTTCGAAAGTCTAAT-----AACGATCAACTACAT 393

Qy 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIle 63
   |||  |||  |||  :::  |||  |||  |||  |||  |||  |||
Db 394 GTAACATCCACTCTTTTAAACGACCTATATTAGATAAGTTACTTAAACACT----- 447

Qy 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyr----- 78
   |||  |||  |||  :::  |||||  |||  |||  |||  |||  |||
Db 448 -----AAAGAAACATTTATCTATACATGGATTCTCTGGAGATGACCTATT 492

Qy 79 -----LysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeu 93
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 493 GTTTATATCGGAGTAAAGATAAAGAGATGTCTCACTCTATCACGAAAGAAATTACGTAAA 552

Qy 94 GluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnPro 113
   ::|||:::  ::|||:::  |||  |||  |||  |||  |||  |||  |||  |||
Db 553 AAAGACTTTACT-----GTAAAGAAAGTCCT 579

Qy 114 AsnProArgAspIleLysPheAspGluArgProHisLysArgArgLysSerArgSer 133
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 580 AACAAATCGATCGCAAGTCATCAGATAACATACGTAATAA----- 621

Qy 134 LysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAla 153
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 622 -----AATGAATCTAACTCAGGTGTCCAAATTA 648

Qy 154 HisPheThrSerLysArgLysAspThrLysArg-----ArgGlnGluArgHisIle 171
   |||||:::  |||||:::  |||||  |||||  |||||  |||||  |||||  |||||
Db 649 GAACTCAGCATGCTGCGGAAACAGTTTTTCAACATTATAAATTAGATCGACATACA 708

Qy 172 LysGluGluGlnAspLys 177
   :::  :::  |||||
Db 709 CGTAGTCATTCAGATAAA 726
```

Search completed: February 25, 2003, 12:53:56  
Job time : 75 secs



Score: 81.50 Matches: 35  
Percent Similarity: 39.71% Conservative: 19  
Best Local Similarity: 25.74% Mismatches: 52  
Query Match: 8.08% Indels: 31  
DB: 3 Gaps: 4

US-09-868-352-23 (1-192) x US-08-680-897-1 (1-9880)

QY 67 TrpGlyAsnGluGlnLeuArgLeuGlyPheTyrlYsAspAlaSerThr----- 83  
Db 8138 TGGAAACAGCGCTCCAGAGTAGAGGAGTCTCGGACACGCGACAGCGGAAG 8197  
QY 84 IleArgLysAsnSerArgIleSerArgLeuGluAspTyrlLysGluTyrlCysAsnPhe 103  
Db 8198 ATACGGAAGCAGAGGAGACAGAGG---AAGAGGCACCGAAGACATGTCAGTGCC 8254  
QY 104 GlyCysAlaTyrlPheValLeuGlu----- 111  
Db 8255 GGGTGTGCGGCGAATATATCAGGCGCATCAGGAGCCCATCTCCAGACCCATGATATGA 8314  
QY 112 -----AsnProAsnProArgAspLys 119  
Db 8315 CGATTCAGGAGTACCGGACAGTACCGTGAGGATGTCCCTTCGCGCGGATGATAAA 8374  
QY 120 PheAspAspGluArgProHisLysArgLysSerArgSerLysSerGlnSerLys 139  
Db 8375 CATGACGAACGAGTCCCGT---AAACGCAAGGTACGAGGTCGAGATCCGCTCTCAGC 8432  
QY 140 SerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLysLys 159  
Db 8433 GCGGACAAAGAAACAGGCGCGATACAGACGCGTGATGATAGGAGATCGAGCGTCTCGAG 8492  
QY 160 ArgLysAspThrLysArgArgGlnGlu-----ArgHisLysLysGlu 173  
Db 8493 AGGCAGAACGAGCGTCTGGCCAGGAGTCTGCTACCCACGTCGAGCAG 8540

RESULT 37  
US-08-961-527-167/c  
; Sequence 167, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 167:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9711 base pairs

; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-167

Alignment Scores:

Pred. No.: 28.8 Length: 9711  
Score: 81.00 Matches: 35  
Percent Similarity: 43.36% Conservative: 27  
Best Local Similarity: 24.48% Mismatches: 47  
Query Match: 8.03% Indels: 34  
DB: 4 Gaps: 8

US-09-868-352-23 (1-192) x US-08-961-527-167 (1-9711)

QY 60 TyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrlYs 79  
Db 1214 TTTAACTATCTTGCCTTGGACTATGGACTCAAAACAAGTCGCAATCTCAGGACTTTCTCCA 1155  
QY 80 AspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrlLysGlu 99  
Db 1154 GATCCAGAGCCA-----TCAGCTGCTCGCTTGCAGNATTGACAGAGTATGTCAAGAAA 1101  
QY 100 TyrCysAsnPheGlyCysAlaTyrlPheValLeuGluAsnProAsnProArgAspLys 119  
Db 1100 AAT---AAAATCGCCTATATCTATTT----- 1077  
QY 120 PheAspAspGluArgProHisLys-----ArgArgLysSerArgSer 133  
Db 1076 -----GAAGAAATGCCTCAACAGCCCTTGTAAACACTTTCAAAAGACGAGTGTGCA 1023  
QY 134 LysSerGlnSerLysSer-----GlnThrArgAsnAsnArgSerGlnSer 149  
Db 1022 AAACGTGATGCTCAATCCTTTAGAAAGTCTGACAGAGAGCACCAGGCTGGAGAAA 963  
QY 150 AsnAlaAsnAla-----HisPheThrSerLys-----LysArgLysAspThrLysArg 165  
Db 962 ACTACATTTTCGTGATGGAGAAAAACCTCAAGGCTTTGAAACAAACACACAGCAAGAAG 903  
QY 166 ArgGlnGluArgHisLysLysGluGlu-----GlnAspLysGluMetThrSer 181  
Db 902 GCCCAGCAATTGAACCTGAAAGGAGGAGGATACCAAGACAGATCCAAAATGGTTACTTCG 843  
QY 182 AlalysGln 184  
Db 842 AGGATGCAG 834

RESULT 38  
US-08-691-563C-46  
; Sequence 46, Application US/08691563C  
; Patent No. 6001987  
; GENERAL INFORMATION:  
; APPLICANT: Herve PERRON  
; APPLICANT: Frederic BESEME  
; APPLICANT: Frederic BEDIN  
; APPLICANT: Glauca PARANHOS-BACCALA  
; APPLICANT: Florence KOMURIAN-PRADEL  
; APPLICANT: Colette JOLIVET  
; APPLICANT: Bernard MANDRAND  
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL

; NUMBER OF SEQUENCES: 92  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Oliff & Berridge  
; STREET: 700 South Washington Street, Suite 300  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22314  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible







```
; LOCATION: 13149..14171
US-09-299-268-14

Alignment Scores:
Pred. No.: 32.6 Length: 14176
Score: 82.50 Matches: 35
Percent Similarity: 41.67% Conservative: 25
Best Local Similarity: 24.31% Mismatches: 45
Query Match: 8.18% Indels: 39
DB: 4 Gaps: 8

US-09-868-352-23 (1-192) x US-09-299-268-14 (1-14176)
QY 5 ValThrProGluMetLeuAsn-----TyrAsnLysTyrProGlyProGln 19
Db 9796 GTTGACCTAATCATGAATGTGGCATCTATTATTATAGCGAATAC-----GAA 9846

QY 20 PheIleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuVal----- 37
Db 9847 TATCTTTTAAAAAATCAAAATTATAAAATAGACAGATAAATTATATCTATATAGATAAG 9906

QY 38 IleAsnGluLysSerAlaPheAspValThrValPhe----- 49
Db 9907 ATTAATAATATCATAGTATAGATATATATTTATCTATCTTCATTTGGAGAAAAAA 9966

QY 50 -----GlyGlnArgPheSerGluIleLeuLysTyrAsp--- 61
Db 9967 TATAACAACACATGCGCATGTGTAGTTATTATTAAGGAA---CTCATGAATATGATATA 10023

QY 62 -----PheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPhe 77
Db 10024 TTAGTACAAATATATATATATATATATATATAAATACA----- 10062

QY 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
Db 10063 TACAAAGAGGGAGATACTATATCCATTAACATACGTTTAAATGTAAGATGATATAATT 10122

QY 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAsp 117
Db 10123 AAA---CATTTGAAGTCTCTTAGTATGTTTGTCTATATATATCATCGAAAAAATCGAC 10179

QY 118 IleLysPheAsp 121
Db 10180 GTAGATTTTGAT 10191

RESULT 32
US-09-866-912-1
; Sequence 1, Application US/09086912
; Patent No. 6166192
; GENERAL INFORMATION:
; APPLICANT: Bruce M. Spiegelman, Pere Puigserver and Zhidan Wu
; TITLE OF INVENTION: PGC-1, A No. 6166192el Brown Pat PPAR(SYMBOL
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/086,912
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,107
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Mandragouras, Amy B.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3066 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 92..2482
US-09-086-912-1

Alignment Scores:
Pred. No.: 4.15 Length: 3066
Score: 82.00 Matches: 38
Percent Similarity: 37.57% Conservative: 30
Best Local Similarity: 20.99% Mismatches: 49
Query Match: 8.13% Indels: 64
DB: 4 Gaps: 8

US-09-868-352-23 (1-192) x US-09-086-912-1 (1-3066)
QY 65 GlyAspTrpGlyAsnGluGln-----LeuArg 73
Db 1547 GCGGAGCTTCAGTAATGACAAATCTCCAACTACCTGTGTTTATAAATTCAGGACTAGCC 1606

QY 74 LeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeu 93
Db 1607 ATGGATGCGCTATTTCATGCACAGTGAAGAT-----GAAAGTGATAAAGTACCTACCT 1660

QY 94 GluAspTyrIleLysGluTyrCysAsnPhe-----GlyCysAlaTyrPhe--- 108
Db 1661 TGGATGGCAGCAGCAGCCCTATTTCATTTTCGATGTCGCGCTTCTTCCTCTTAAAC 1720

QY 109 -----ValLeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgProHis 126
Db 1721 TCTCCGTGTCGAGACACAGTGTGTCACCCGAAATCTTTATTT---TCTCAAGACCCCAA 1777

QY 127 LysArgArg----- 129
Db 1778 AGGATGCGCTCTGTTCAAGATCCTTTTCTCGACACAGGTGCGTGTCCCGATCACCATAT 1837

QY 130 ---LysSerArgSerLysSerGlnSerLysSerGlnThrArgAsn----- 144
Db 1838 TCCAGGTCAAGATCAAGGTCCCGAGCAGTAGATCCTCTTCAAGATCTGTACTACTAT 1897

QY 145 -----AsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLys 158
Db 1898 GAATCAAGCCACTACAGACACCCGACACCCGCAATCTCCCTGTGTGTGAGATCAGT 1957

QY 159 LysArgLysAspThrLysArg----- 166
Db 1958 TCAAGGTCAACCTACAGCCGTAGGCCCGAGGTACGACAGCTATGAAGCCTATGAGCAGAA 2017

QY 167 -----GlnGluArgHisIleLysGluGluAsnLysGluMetThrSerAlaLys 193
Db 2018 AGGCTCAAGAGGGATGAATACCCAAAGAGCAGCAGAGCGGAGGAGTCTCAAAGGGCCAA 2077

QY 184 Gln 184
Db 2078 CAG 2080

RESULT 33
US-09-203-453-1
; Sequence 1, Application US/09203453
; Patent No. 6426411
; GENERAL INFORMATION:
```



```

; LOCATION: 7454..7858
; FEATURE: NAME/KEY: CDS
; LOCATION: 7895..8155
; FEATURE: NAME/KEY: CDS
; LOCATION: 8215..8682
; FEATURE: NAME/KEY: CDS
; LOCATION: 8715..9539
; FEATURE: NAME/KEY: CDS
; LOCATION: 9562..10272
; FEATURE: NAME/KEY: CDS
; LOCATION: 10316..11908
; FEATURE: NAME/KEY: CDS
; LOCATION: 11971..12780
; FEATURE: NAME/KEY: CDS
; LOCATION: 12829..13107
; FEATURE: NAME/KEY: CDS
; LOCATION: 13149..14171
; US-08-307-499-14

Alignment Scores:
Pred. No.: 32.6 Length: 14176
Score: 82.50 Matches: 35
Percent Similarity: 41.67% Conservatives: 25
Best Local Similarity: 24.31% Mismatches: 45
Query Match: 8.18% Indels: 39
DB: 1 Gaps: 8

US-09-868-352-23 (1-192) x US-08-307-499-14 (1-14176)
QY 5 ValThrProGluMetLeuAsn-----TyrAsnLysTyrProGlyProGln 19
Db 9796 GTTGTACCTAATCATAGTAAGTGGCATCTATTATTTATAGCGAATAC-----GAA 9846
QY 20 PheIleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuVal----- 37
Db 9847 TATCTTTTAAAAAATCAAAATTTATAAAATAAGAAATAAATATATATATAGATAAG 9906
QY 38 IleAsnGluLysSerAlaPheAspValThrValPhe----- 49
Db 9907 ATTAATAAATATCATAGTAGATGATATATATATATATATATCTTCTTCATGAGAAAAA 9966
QY 50 -----GlyGlnArgPheSerGluIleLeuLysTyrAsp--- 61
Db 9967 TATAACAACACATGCCGATGTGGTAAGTTATTTAAGGAA---CTCATGAATATGATATA 10023
QY 62 -----PheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPhe 77
Db 10024 TTAGCTACAAAATATATATATATATATATATATATAAATACA----- 10062
QY 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
Db 10063 TACAAGAGGGAGATCATATATCATTAACATACGTTTAAATGTAAGATGATATAATT 10122
QY 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAsp 117
Db 10123 AAA---CATTGTAAGTCTTCATAGGTATGTTGCTATATATATATATATATATATCGAC 10179
QY 118 IleLysPheAsp 121
Db 10180 GTAGATTTTCAT 10191

RESULT 30
US-09-299-268-1/c
; Sequence 1, Application US/09299268
; Patent No. 6217882

; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Foxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,268
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/901,127
; FILING DATE:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14176 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3852..4226
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4585..4887
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5131..5310
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5760..5912
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6786..7130
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10148..10513
; US-09-299-268-1

Alignment Scores:
Pred. No.: 32.6 Length: 14176
Score: 82.50 Matches: 35
```

```
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3852..4226
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4585..4887
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5131..5310
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5760..5912
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6786..7130
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10148..10513
; US-08-307-499-1
;
; Alignment Scores:
; Pred. NO.: 32.6
; Score: 82.50
; Percent Similarity: 41.67%
; Best Local Similarity: 24.31%
; Query Match: 8.18%
; DB: 1
;
; US-09-868-352-23 (1-192) x US-08-307-499-1 (1-14176)
;
; Qy 5 ValThrProGluMetLeuAsn-----TyrAsnLysTyrProGlyProGln 19
; ||| ||| :||| |||:||||| :||
; Db 4381 GTTGACCTAATCATATGAATGCGATCTATTATTTATAGCAATAC-----GAA 4331
;
; Qy 20 PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuVal----- 37
; :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
; Db 4330 TATCTTTTAAAAAATCAATATAAATAAAGAGATAAATATATCTATATTAGATAAG 4271
;
; Qy 38 IleAsnGluLysSerAlaPheAspValThrValPhe----- 49
; |||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
; Db 4270 ATTAATAAATATCATAGTATAGATATATATTTATGTATCTTCATTGCGAGAAAAA 4211
;
; Qy 50 -----GlyClnArgPheSerGluIleLeuLeuLysTyrAsp--- 61
; |||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
; Db 4210 TATAACAACATGCGCATGCGTAAAGTTATTAAAGAA---CTCATGAAATATGATATA 4154
;
; Qy 62 -----PheIleValGlyAspTrpGlyAsnGluGlnLeuAtrGlyPhe 77
; :|||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
; Db 4153 TTACGTACAAAATATATATAATGATATTTATAATACA----- 4115
;
; Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
; |||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
; Db 4114 TACAAAGAGGAGATACTATATCCATTAAACATACGTTTAAATGTAAGATGATATAAT 4055
;
; Qy 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAsp 117
; ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
; Db 4054 AAA---CATCTAAGCTCTCTATAGGTATGTTGCTATATTATCATCGAAAAAATATCGAC 3998
;
; Qy 118 IleLysPheAsp 121
; :|||:|||||
; Db 3997 GTAGATTTTGAT 3986
;
; RESULT 29
; US-08-307-499-14
; Sequence 14, Application US/08307499
; Patent No. 5651972
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
```

```
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/307,499
; APPLICATION NUMBER: US/08/307,499
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: OF35.1.PWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14176 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 138..1460
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2456..2659
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2809..3030
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3070..3330
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3356..4180
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4392..5894
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6171..6398
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6447..6875
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6928..7431
; FEATURE:
; NAME/KEY: CDS
```

Db 817 TCATCATCTAAAGACACCAAGTTAGAAAGAGAAACTACAGTCAAGAGGAATCT 876  
QY 176 AsPLysGluMetThrSerAlaLysGln 184  
Db 877 AAAGCCACAGCACTAAAGTGAAGAA 903  
RESULT 27  
US-09-402-328-1/c  
; Sequence 1, Application US/09402328  
; Patent No. 6365728  
; GENERAL INFORMATION:  
; APPLICANT: Purdue Research Foundation,  
; Hodges, Thomas K.  
; Lysnik, Leszek A.  
; TITLE OF INVENTION: Regulatory Element For Expressing Genes  
; In Plants  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Barnes & Thornburg  
; STREET: 11 S. Meridian  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/09/402,328  
; FILING DATE: 05-No. 6365728-1999  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Breen, John P.  
; REGISTRATION NUMBER: 38,833  
; REFERENCE/DOCKET NUMBER: 3220-29933  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317) 231-7745  
; TELEFAX: (317) 231-7433  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5285 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Arabidopsis thaliana  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-402-328-1  
Alignment Scores:  
Pred. No.: 7.89 Length: 5285  
Score: 82.50 Matches: 37  
Percent Similarity: 39.33% Conservative: 22  
Best Local Similarity: 24.67% Mismatches: 52  
Query Match: 8.18% Indels: 39  
DB: 4 Gaps: 6  
US-09-868-352-23 (1-192) x US-09-402-328-1 (1-5285)  
QY 34 PheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlnArgPhe 53  
Db 2372 TATCTTTATTTTCAGATATAAAAGCTCGATGACACAGAGTTTCGGAATCGGTTTT 2313  
QY 54 SerGluIleLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluLeuArg 73  
Db 2312 -----CTCATGAAT 2304

QY 74 LeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeu 93  
Db 2303 ATGTCGGGTATCTTAAGAC-----AAGAATTCT----- 2274  
QY 94 GluAspTyrIleLysGluTyrCys-----AsnPheGlyCysAlaTyrPheValLeuGlu 111  
Db 2273 GAGAGCTGCATTCGTTTATTGCACCTTAACATTTCCCTATGTCATTACAGGTTAGAGAA 2214  
QY 112 AsnProAsnProArgAspIleLys-----PheAspAspGluArg 124  
Db 2213 TATGATTCAAGAGGAGGATTCGAGGAGCCCGCGTGAAGATCCTATTCTAGAGCCGC 2154  
QY 125 ProHisLysArgLysSerArgSerLysSerGlnSer-----SerLysSerGln 141  
Db 2153 AGCCGTGGACGAGTCCCGCTAGTCGTAGCCGACGAGCAAGAGCAGAGCAGAGA 2094  
QY 142 ThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLys 161  
Db 2093 AGTCCAAAGGCTAAATCTTTGCGTAGATCGCGTCAAAATCTACATCGAGATCTCCTCGC 2034  
QY 162 AspThrLysArgArgGlnGluArgHisIle 171  
Db 2033 TCTCGCTCCCGCTCTAAGTCGAGGTCACTG 2004  
RESULT 28  
US-08-307-499-1/c  
; Sequence 1, Application US/08307499  
; Patent No. 5651972  
; GENERAL INFORMATION:  
; APPLICANT: Moyer, Richard W.  
; APPLICANT: Vi uela, Eladio  
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: U.S.A.  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/307,499  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/908,241  
; FILING DATE: 1-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/908,630  
; FILING DATE: 29-JUN-1992  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/342,212  
; FILING DATE: 21-APR-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14176 base pairs  
; TYPE: nucleic acid

```
Alignment Scores:
Pred. No.: 3.74 Length: 3143
Score: 82.50 Matches: 42
Percent Similarity: 36.51% Conservative: 27
Best Local Similarity: 22.22% Mismatches: 47
Query Match: 8.18% Indels: 73
DB: 2 Gaps: 9

US-09-868-352-23 (1-192) x US-08-973-831-1 (1-3143)
Qy 14 LysTyrProGlyProGlnPheIleHisPheGluAsnIleVal-----LysSerAsp 30
Db 502 CGTTATTTAAATCCATTATTGTTGTCATATAGCATATTCCTATGAGTTTCATCAGAC 561
Qy 31 AspileGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50
Db 562 GAAGAACACATTGCTTCTAGACTCCTCTAAATAATCATCTCGATCACT----- 612
Qy 51 GlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70
Db 612 ----- 612
Qy 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr----- 83
Db 613 -----TCAGCTTCCACTTATGAAGACGATCAAGATGAT 645
Qy 84 -----IleArgLysAsnSerArgIleSerArgLeuGlu---AspTyrIleLysGlu 99
Db 646 GATATCCCTTTAGCTAAATAATCCAGGAAAGAGGGTTGAATCTGTATGTAAGAGAGAT 705
Qy 100 TyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspileLys 119
Db 705 ----- 705
Qy 120 PheAspAspGluArgProHisLysArgArgLys-----SerArgSerLysSerGln 136
Db 706 ---GAAGACGAAGTCCCATTCATTGAAAGAGAGAAATTTGCTAATGCTAGACGAGCAAAACAA 762
Qy 137 SerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAlaHisPheThr 156
Db 763 ---GTTAAACCGAACTAAAGTTAAAGAGAACCTAAAGTGCCATTAATATCC---AAA 816
Qy 176 AspileGluMetThrSerAlaLysGln 184
Db 877 AAAGCCACAGCACATAAGTGAAGAGAA 903

RESULT 26
PCT-US96-09530A-1
; Sequence 1, Application PC/TUS9609530A
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gerhold, David L.
; APPLICANT: Strauss, Allison Cole
; TITLE OF INVENTION: Anti-fungal Agents and Methods of
; TITLE OF INVENTION: Identifying and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.1
; CURRENT APPLICATION DATA:
```



```

; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-307-499-49

Alignment Scores:
Pred. No.: 0.442 Length: 711
Score: 82.50 Matches: 35
Percent Similarity: 41.67% Conservative: 25
Best Local Similarity: 24.31% Mismatches: 45
Query Match: 8.18% Indels: 39
DB: 1 Gaps: 8

US-09-868-352-23 (1-192) x US-08-307-499-49 (1-711)
Qy 5 ValThrProGluMetLeuAsn-----TyrAsnLysTyrProGlyProGln 19
Db 235 GTTGTTACCTAATCATATGAATGTGCATCTATTATTATAGCGAATAC-----GAA 285
Qy 20 PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuVal----- 37
Db 286 TATCTTTTAAAAAATCAAAATTAATAAATAAGAGATAAATTATATATAGATAAG 345
Qy 38 IleAsnGluLysSerAlaPheAspValThrValPhe----- 49
Db 346 ATTAATAAATCATAGTAGATGATATATATTTATGTATCTTCATTGGAGAAAAA 405
Qy 50 -----GlyClnArgPheSerGluIleLeuLeuLysTyrAsp--- 61
Db 406 TATAACAACACATGCGCATGTGGTAAGTTATTAAAGGAA---CTCATGAAATATGATATA 462
Qy 62 -----PheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPhe 77
Db 619 GTAGATTTTGAT 630
Qy 118 IleLysPheAsp 121
Db 619 GTAGATTTTGAT 630

RESULT 23
US-09-299-268-49
; Sequence 49, Application US/09299268
; Patent No. 6217882
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,268
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/901,127
; FILING DATE:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCCI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-09-299-268-49

Alignment Scores:
Pred. No.: 0.442 Length: 711
Score: 82.50 Matches: 35
Percent Similarity: 41.67% Conservative: 25
Best Local Similarity: 24.31% Mismatches: 45
Query Match: 8.18% Indels: 39
DB: 1 Gaps: 8

US-09-868-352-23 (1-192) x US-09-299-268-49 (1-711)
Qy 5 ValThrProGluMetLeuAsn-----TyrAsnLysTyrProGlyProGln 19
Db 235 GTTGTTACCTAATCATATGAATGTGCATCTATTATTATAGCGAATAC-----GAA 285
Qy 20 PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuVal----- 37
Db 286 TATCTTTTAAAAAATCAAAATTAATAAATAAGAGATAAATTATATATAGATAAG 345
Qy 38 IleAsnGluLysSerAlaPheAspValThrValPhe----- 49
Db 346 ATTAATAAATCATAGTAGATGATATATTTATGTATCTTCATTGGAGAAAAA 405
Qy 50 -----GlyClnArgPheSerGluIleLeuLeuLysTyrAsp--- 61
Db 406 TATAACAACACATGCGCATGTGGTAAGTTATTAAAGGAA---CTCATGAAATATGATATA 462
Qy 62 -----PheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPhe 77
Db 463 TTAGCTACAAAATATATATATAATGATATATAATAACA----- 501
Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
Db 502 TACAAAGAGGGAGAGACTACTATATCCATTAAACATACGTTTAAAAATGTAAGATGATATA 561
Qy 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgasp 117
Db 562 AAA---CATTTGAAGTCTTCTATAGGTATGTTGCTGCTATATTATATATATATATATCGAC 618
Qy 118 IleLysPheAsp 121
Db 619 GTAGATTTTGAT 630

RESULT 24

```







; FILING DATE: 19-AUG-1993  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Connell, Gary J.  
 ; REGISTRATION NUMBER: 32,020  
 ; REFERENCE/DOCKET NUMBER: 2618-13  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 303/863-9700  
 ; TELEFAX: 303/863-0223  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 913 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 3..911  
 ; US-08-109-391A-1

Alignment Scores:  
 Pred. No.: 0.551 Length: 913  
 Score: 83.00 Matches: 41  
 Percent Similarity: 42.21% Conservative: 24  
 Best Local Similarity: 26.62% Mismatches: 42  
 Query Match: 8.23% Indels: 47  
 DB: 1 Gaps: 9

US-09-868-352-23 (1-192) x US-08-109-391A-1 (1-913)

QY 68 GlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArg---Lys 86  
 DB 456 GGTAAGAAATGATCCCAAAATGATCGATGTAATCGAAATATGATGATGATGAA 506  
 QY 87 AsnSerArg---IleSerArgLeuGluAspTyrIleLysGluTyr---CysAsnPheGly 104  
 DB 507 AATGGAATATGATCCCAAAATGATCGATGTAATCGAAATATGATGATGATGATG 566  
 QY 105 -----CysAlaTyrPheValLeu----- 110  
 DB 567 ACAGATGAACAACATTTGTAATATTTGTCGAAGCTTTGCAACAGCGAGAGGTGTAACG 626  
 QY 111 -----GluAsnProAsnProArg----- 116  
 DB 627 GTGCAGGATAATGCAATTCGAGATGACGAGATACCAAAATTTGAGGCGAGGAGGAAGAGAA 686  
 QY 117 -----AspIleLysPheAspGluArg-ProHisLysArgArgLysSerArgSe 133  
 DB 687 AAATACGATCAAGTAAAGGAAGATAGGAGCGCGAATCGAAGAGGAGGAGGAACAGAA 746  
 QY 133 rLysSerGlnSerSerLysSerGlnThrArgAsnAsn-----ArgSerGlnSerAs 150  
 DB 747 AGGCTGAGAGAGTACGAGGACAGATACAGAAAAATTTGAGGCGAGGAGGAAGAGAGAA 806  
 QY 150 nAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgH 170  
 DB 807 C-----GGCAAGAACAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATG 848  
 QY 170 sIleLysGluGlnAspLysGluMetThrSerAlaLys 183  
 DB 849 GAACAAGAGGTAACACAAAGATATGATGAAAGGAAA 888

RESULT 18  
 US-08-459-019A-1  
 ; Sequence 1, Application US/08459019A  
 ; Patent No. 5686080  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tripp, Cynthia A.  
 ; APPLICANT: Frank, Glenn R.  
 ; APPLICANT: Grievie, Robert B.  
 ; TITLE OF INVENTION: NOVEL PARASITIC HELMINTH P4 PROTEINS  
 ; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sheridan Ross & McIntosh  
 ; STREET: 1700 Lincoln Street, #3500  
 ; CITY: Denver  
 ; STATE: CO  
 ; COUNTRY: U.S.A.  
 ; ZIP: 80203  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/459,019A  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Connell, Gary J.  
 ; REGISTRATION NUMBER: 32,020  
 ; REFERENCE/DOCKET NUMBER: 2618-13-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (303) 863-9700  
 ; TELEFAX: (303) 863-0223  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 913 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Dirofilaria immitis  
 ; DEVELOPMENTAL STAGE: Larva  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: L3 and/or L4 larval D. immitis cDNA  
 ; CLONE: p4  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 3..911  
 ; US-08-459-019A-1

Alignment Scores:  
 Pred. No.: 0.551 Length: 913  
 Score: 83.00 Matches: 41  
 Percent Similarity: 42.21% Conservative: 24  
 Best Local Similarity: 26.62% Mismatches: 42  
 Query Match: 8.23% Indels: 47  
 DB: 1 Gaps: 9

US-09-868-352-23 (1-192) x US-08-459-019A-1 (1-913)

QY 68 GlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArg---Lys 86  
 DB 456 GGTAAGAAATGATCCCAAAATGATCGATGTAATCGAAATATGATGATGATGAA 506  
 QY 87 AsnSerArg---IleSerArgLeuGluAspTyrIleLysGluTyr---CysAsnPheGly 104  
 DB 507 AATGGAATATGATCCCAAAATGATCGATGTAATCGAAATATGATGATGATGATG 566  
 QY 105 -----CysAlaTyrPheValLeu----- 110  
 DB 567 ACAGATGAACAACATTTGTAATATTTGTCGAAGCTTTGCAACAGCGAGAGGTGTAACG 626  
 QY 111 -----GluAsnProAsnProArg----- 116  
 DB 627 GTGCAGGATAATGCAATTCGAGATGACGAGATACCAAAATTTACTGTATCCATGGAACAG 686  
 QY 117 -----AspIleLysPheAspGluArg-ProHisLysArgArgLysSerArgSe 133  
 DB 687 AAATACGATCAAGTAAAGGAAGATAGGAGCGCGAATCGAAGAGGAGGAGGAACAGAA 746

```

CORRESPONDENCE ADDRESS:
ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/623.906A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J.
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: A-62282/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 454 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..133
OTHER INFORMATION: /note= "Nucleotides 1-133 are
OTHER INFORMATION: unique flanking sequence"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 134..307
OTHER INFORMATION: /note= "Nucleotides 134-207 are
OTHER INFORMATION: repeat sequence"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 308..454
OTHER INFORMATION: /note= "Nucleotides 308-454 are
OTHER INFORMATION: unique flanking sequence"
US-08-623-906A-6

Alignment Scores:
Pred. No.: 0.175 Length: 454
Score: 83.50 Matches: 17
Percent Similarity: 50.77% Conservative: 16
Best Local Similarity: 26.15% Mismatches: 25
Query Match: 8.28% Indels: 7
DB: 2 Gaps: 1

US-09-868-352-23 (1-192) x US-08-623-906A-6 (1-454)
Qy 119 LysPheAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSer 138
||| : : : : : ||| : : : : : ||| : : : : :
135 AAGAAAGAAAAAAGAACCAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAA 194
||| : : : : : ||| : : : : : ||| : : : : :
139 LysSerGlnThrArgAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLys 158
||| : : : : : ||| : : : : : ||| : : : : :
195 AAGAAAGAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 233
||| : : : : : ||| : : : : : ||| : : : : :
159 LysArgLysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGlu 178
||| : : : : : ||| : : : : : ||| : : : : :
234 AAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 293
||| : : : : : ||| : : : : : ||| : : : : :
Qy 179 MetThrSerAlaLys 183
|||
294 AGAAAGAAAGAAAGAAAG 308
|||

```

```

RESULT 16
US-09-146-053-5
: Sequence 5, Application US/09146053A
: Patent No. 639349
: GENERAL INFORMATION:
: APPLICANT: Ryan, James W.
: APPLICANT: Sprinkle, Terry Joe Curtis
: APPLICANT: Venema, Richard C.
: TITLE OF INVENTION: Human Aminopeptidase P Gene
: FILE REFERENCE: MCG103
: CURRENT APPLICATION NUMBER: US/09/146.053A
: CURRENT FILING DATE: 1998-09-02
: EARLIER APPLICATION NUMBER: 60/057,854
: EARLIER FILING DATE: 1997-09-02
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 44453
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-146-053-5

Alignment Scores:
Pred. No.: 127 Length: 44453
Score: 83.50 Matches: 18
Percent Similarity: 49.21% Conservative: 13
Best Local Similarity: 28.57% Mismatches: 31
Query Match: 8.28% Indels: 1
DB: 4 Gaps: 1

US-09-868-352-23 (1-192) x US-09-146-053-5 (1-44453)
Qy 123 GluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThr 142
||||| : : : : : ||| : : : : : ||| : : : : :
Db 20929 GAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 20988
||| : : : : : ||| : : : : : ||| : : : : :
Qy 143 ArgAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAsp 162
||| : : : : : ||| : : : : : ||| : : : : :
Db 20989 AGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 21048
||| : : : : : ||| : : : : : ||| : : : : :
Qy 163 Thr---LysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSer 181
||| : : : : : ||| : : : : : ||| : : : : :
Db 21049 ACAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 21108
||| : : : : : ||| : : : : : ||| : : : : :
Qy 182 AlaLysGln 184
||| : : : : :
Db 21109 AGAAAGAAAGAA 21117

RESULT 17
US-08-109-391A-1
: Sequence 1, Application US/08109391A
: Patent No. 5639876
: GENERAL INFORMATION:
: APPLICANT: Tripp, Cynthia A.
: APPLICANT: Frank, Glenn R.
: APPLICANT: Grieve, Robert B.
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NOVEL
: TITLE OF INVENTION: PARASITIC HELMINTH PROTEINS
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheridan Ross & McIntosh
: STREET: 1700 Lincoln St., Suite 3500
: CITY: Denver
: STATE: CO
: COUNTRY: U.S.A.
: ZIP: 80203
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/109,391A

```



[illegible]



```
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Dictyostelium discoideum
US-08-914-999-7

Alignment Scores:
Pred. No.: 0.652 Length: 2237
Score: 87.00 Matches: 45
Percent Similarity: 34.42% Conservative: 29
Best Local Similarity: 20.93% Mismatches: 69
Query Match: 8.62% Indels: 72
DB: 4 Gaps: 11

US-09-868-352-23 (1-192) x US-08-914-999-7 (1-2237)

Qy 12 TyrAsnLysTyrProGly-----ProGln---PheIle 21
Db 739 TATAATAATAATAGTGGATTCGTTAGTAATGATGAAGAAATACACCAATCATCTCT 798
Qy 22 HisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLys 41
Db 799 CATTC-----ACATATGAACATTCAAATCATCAATTATTCATT----- 837
Qy 42 SerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeu---LysTyr 60
Db 838 -----ATAGATATTCAAGGTGTTGGTGATCACTATACAGACCCACAATTCATACCTAT 891
Qy 61 Asp-----PheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPhe 77
Db 892 GATGCTGTTGGTTTGGTATTGGTAATTTGGTCAA-----AAAGGTTTT 936
Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
Db 937 GAAAGTTTTATAGATCACTATAA----- 960
Qy 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArg--- 116
Db 961 -----TGATATGCAATTGGCAATATTAAATTTACATCAATTAATCCAAATCT 1011
Qy 117 -----AspIleLysPheAspGluArgPro 125
Db 1012 GAAAAAGTGTCTGGTACTGTACCAAGACAGATTAAATTTCCCTGATACATCTGAA 1071
Qy 126 HisLysArgArgLysSerArgSerLysSerGlnSerLysSerGlnThrArgAsnAsn 145
Db 1072 ACAGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1131
Qy 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLys-----Arg 160
Db 1132 AATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1191
Qy 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGln----- 175
Db 1192 AGTGCTAGTAAAGAAAGAAATGATAGATTCGCAAGTAGACATTAATTTGTTTCAAT 1251
Qy 176 -----AspLysGluMetThrSerAlaLys 183
Db 1252 GATGCTAATACATTAAATACAAATAAAGAGATCAAAATCAAAA 1296

RESULT 8
US-07-922-723A-7
; Sequence 7, Application US/07922723A
; Patent No. 5369004
; GENERAL INFORMATION:
; APPLICANT: Drs. Michael H. Polymeropoulos
; APPLICANT: and Carl R. Merril
; TITLE OF INVENTION: FIVE HIGHLY INFORMATIVE
; TITLE OF INVENTION: REPEAT POLYMORPHIC DNA MARKERS
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker

; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/922.723A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: D.J. Mills
; REGISTRATION NUMBER: 34506
; REFERENCE/DOCKET NUMBER: 717081B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-922-723A-7

Alignment Scores:
Pred. No.: 0.046 Length: 291
Score: 86.00 Matches: 24
Percent Similarity: 36.70% Conservative: 16
Best Local Similarity: 22.02% Mismatches: 37
Query Match: 8.52% Indels: 32
DB: 1 Gaps: 2

US-09-868-352-23 (1-192) x US-07-922-723A-7 (1-291)

Qy 79 LysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLys 98
Db 43 AAAGACAAAGAGAGATTAGAAAGAAAGAAAGAA----- 75
Qy 99 GluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle 118
Db 76 -----GAGAGAGAGAGAAAGAGAGAGG 99
Qy 119 LysPheAspAspGluArgProHisLysArgLysSerArgSerLysSerGlnSerSer 138
Db 100 AAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 159
Qy 139 LysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLys 158
Db 160 AAGAAAGAAAGAAAGAAAGAAAGAAAGAA-----AAG 186
Qy 159 LysArgLysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGlu 178
Db 187 AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 246
Qy 179 MetThrSerAlaLysGlnHisLeuLeu 187
Db 247 AGGAAGGAAAGAAAGAGCAAGTTACTA 273

RESULT 9
US-07-799-828C-7
; Sequence 7, Application US/07799828C
; Patent No. 5378602
; GENERAL INFORMATION:
; APPLICANT: Drs. Carl R. Merril and
; APPLICANT: Michael H. Polymeropoulos
; TITLE OF INVENTION: TWENTY SEVEN HIGHLY INFORMATIVE
; TITLE OF INVENTION: MICROSATELLITE REPEAT
; TITLE OF INVENTION: POLYMORPHIC DNA MARKERS
```



```
Best Local Similarity: 24.29% Mismatches: 24
Query Match: 4 Indels: 9
DB: Gaps: 1
US-09-868-352-23 (1-192) x US-09-018-584A-27 (1-333)
QY 123 GluArgProHisLysArgLysSerArgLysSerGlnSerSerLysSerGlnThr 142
   :: ||| |||::: ::: ||| ::: ||| :::
Db 43 CAGCTCCCAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 102
QY 143 ArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAsp 162
   ||| ::: ||| ::: ||| ::: ||| ::: ||| ::: ||| ::: ||| :::
Db 103 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 135
QY 163 ThrLysArgArgGlnGluArgHisLysLysGluGlnAspLysGluMetThrSerAla 182
   ||| ::: ||| ::: ||| ::: ||| ::: ||| ::: ||| ::: ||| :::
Db 136 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 195
QY 183 LysGlnHisLeuLeuPheValArgLysAsn 192
   ::: ||| ||| ::: ||| |||
Db 196 GAACAGAAACTAAGAAGGCTGAGAGCCAC 225
RESULT 6
US-07-667-276A-3
; Sequence 3, Application US/07667276A
; Patent No. 5470971
; GENERAL INFORMATION:
; APPLICANT: Kondo, Keiji
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING
; TITLE OF INVENTION: THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 S. Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/667,276A
; FILING DATE: 11-MAR-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 377.5351P
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: S288C
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 484..1725
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1
; OTHER INFORMATION: /note= "Base #1 of Sequence No. 5470971 3
; OTHER INFORMATION: corresponds to base -483 of the sequence listed in
; OTHER INFORMATION: Figure 11 of the application"
US-07-667-276A-3
Alignment Scores:
Pred. No.: 0.562 Length: 2017
Score: 87.00 Matches: 19
Percent Similarity: 61.29% Conservative: 19
Best Local Similarity: 30.65% Mismatches: 23
Query Match: 8.62% Indels: 1
DB: Gaps: 0
US-09-868-352-23 (1-192) x US-07-667-276A-3 (1-2017)
QY 124 ArgProHisLysArgArgLysSerArg-SerLysSerGlnSerSerLysSerGlnThrAr 143
   ||| ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 CGAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 836
QY 143 gAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspTh 163
   ::: ||| ||| ::: ||| ::: ||| ::: ||| ::: ||| ::: ||| :::
Db 837 CTCTGATTCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC 896
QY 163 rLysArgArgGlnGluArgHisLysLysGluGlnAspLysGluMetThrSerAlaLy 183
   ||| ::: ||| ::: ||| ::: ||| ::: ||| ::: ||| ::: ||| ::: |||
Db 897 TAAGAAACGTAAATCTCAGGACGCCGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 956
QY 183 sGln 184
   |||
Db 957 GCAA 960
RESULT 7
US-08-914-999-7
; Sequence 7, Application US/08914999
; Patent No. 6346406
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey G.
; APPLICANT: Hait, William N.
; APPLICANT: Pavur, Karen S.
; TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914,999
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 601-1-078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2237 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

DB:	2	Gaps:	2	
US-09-868-352-23 (1-192) x US-08-718-661-1 (1-2790)				
Qy	130	LysSerArgSerLysSerGlnSer---	SerLysSerGluThrArgAsn---AsnArgSer 147	
Db	2295	CAGGCCAGACGAGCAGCCAGACCGAGCCAGACCGAGCCAGACCGGAAACCGGAAACGGAGC	2355	
Qy	148	GlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArgargGln	167	
Db	2355	CAGAGCCAGAGCCAGAACACCAGACCCAGAGGAGACAGCAAGAGCAGACGAGNAGAGGCAG	2414	
Qy	168	GluAArgHisTleLysGluGluInAspLysGluMetThrSerAlaLysGlnHisLeuLeu	187	
Db	2415	AGGAAGGAGCAGAGAAGGAGGAGCAGAACCCAGGACGAGCAGACGAGGAGGAAGAGGAGGAG	2474	
Qy	188	PheValArgLysAsn	192	
Db	2475	AGGAAGCGGAAGAGC	2489	
RESULT 5				
US-09-018-584A-27				
; Sequence 27, Application US/09018584A				
; Patent No. 6238863				
; GENERAL INFORMATION:				
; APPLICANT: Schumm, James W.				
; APPLICANT: Bacher, Jeffery W.				
; TITLE OF INVENTION: MATERIALS AND METHODS FOR				
; TITLE OF INVENTION: IDENTIFYING AND ANALYZING INTERMEDIATE TANDEM				
; TITLE OF INVENTION: REPEAT DNA MARKERS				
; NUMBER OF SEQUENCES: 147				
; CORRESPONDENCE ADDRESS:				
; ADDRESSEE: Promega Corporation				
; STREET: 2800 Woods Hollow Road				
; CITY: Madison				
; STATE: Wisconsin				
; COUNTRY: U.S.A.				
; ZIP: 53711-5399				
; COMPUTER READABLE FORM:				
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb				
; COMPUTER: IBM compatible PC				
; OPERATING SYSTEM: Windows 95				
; SOFTWARE: Word 97 (DOS text format)				
; CURRENT APPLICATION DATA:				
; APPLICATION NUMBER: US/09/018,584A				
; FILING DATE: 04-Feb-1998				
; CLASSIFICATION:				
; ATTORNEY/AGENT INFORMATION:				
; NAME: Grady J. Frenchick				
; REGISTRATION NUMBER: 29,018				
; REFERENCE/DOCKET NUMBER: 16026.9180				
; TELECOMMUNICATION INFORMATION:				
; TELEPHONE: (608) 257-3501				
; TELEFAX: (608) 257-2275				
; INFORMATION FOR SEQ ID NO: 27:				
; SEQUENCE CHARACTERISTICS:				
; LENGTH: 333 bp				
; TYPE: Nucleic Acid				
; STRANDEDNESS: Double				
; TOPOLOGY: Circular				
; MOLECULE TYPE: Genomic DNA				
; HYPOTHETICAL: no				
; IMMEDIATE SOURCE:				
; LIBRARY: plasmid, pGem3zf(+)				
; CLONE: G539				
; POSITION IN GENOME:				
; CHROMOSOME/SEGMENT: 15q26.2				
US-09-018-584A-27				
Alignment Scores:				
Pred. No.:		0.0367	Length:	333
Score:		87.50	Matches:	17
Percent Similarity:		52.86%	Conservative:	20

;  
; LENGTH: 5558 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-103

Alignment Scores:  
Pred. No.: 5,72e-43 Length: 5558  
Score: 438.00 Matches: 88  
Percent Similarity: 68.8% Conservative: 36  
Best Local Similarity: 48.8% Mismatches: 50  
Query Match: 43.41% Indels: 6  
DB: 4 Gaps: 4

US-09-868-352-23 (1-192) x US-08-961-527-103 (1-5558)

QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20  
|||||  
Db 4885 ATCGGAAAGAAATTCACCTGAATATACACTATAACAAGTTCCCTGTCGGAGTTC 4826  
QY 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40  
|||  
Db 4825 CATTTACACGGGACAGAGTCGAAACGGAAGGATAGCTTTTCTTGGTTGAAATATC 4766  
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60  
|||  
Db 4765 AAGGATGCCCTTCATGTGACGACTTTTAATACGCGTTTTCAGAAAGTATTAACCAAGTTT 4706  
QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80  
|||||  
Db 4705 GATTATATCTGGGGACTTGGAGCAACGACGCTTCGCTACGAGGTTTTCACAAAGAT 4646  
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100  
|||  
Db 4645 GACCGAACAAGAGAAATCTGAAATAATCAGTCGTTTACAGACTACCTTTTAGAGTAT 4586  
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120  
|||||  
Db 4585 TGTAGTTATGGTTGTGCTTATTTGCTTAGAAAATGAAGCCCTTAAGCGACATCATTT 4526  
QY 121 AspAspGluArgProHisLysArgLysSerArgLysSerGlnSerLysSer 140  
|||  
Db 4525 GAC-----AAGAAATGCGTAGACGGAGAGAAACACCTTCTAGAAAAGAGAAA 4472  
QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160  
|||  
Db 4471 CGGACTCAAACTAAACGA---AAGTCGAATGCAGATAAGAAA---AATAGACGCTGCTCAG 4418  
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlu-----GlnAspLysGlu 178  
|||||  
Db 4417 AAAGACGACAAGGGACAACGCTATTTTGTGCTATTCGTAGAGTGTAGAGATAAGGAG 4358

RESULT 2  
US-09-134-001C-2057  
; Sequence 2057, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2057  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2057

Alignment Scores:  
Pred. No.: 2,5e-22 Length: 399  
Score: 254.50 Matches: 47  
Percent Similarity: 68.93% Conservative: 24  
Best Local Similarity: 45.63% Mismatches: 29  
Query Match: 25.22% Indels: 3  
DB: 4 Gaps: 2

US-09-868-352-23 (1-192) x US-09-134-001C-2057 (1-399)

QY 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlnLysSerAla 43  
|||||  
Db 4 GAAATATGATTAAAGTCGACCAACAATATTTGAATTGATAGAGAATATAGAGAATGT 63  
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIle 63  
|||||  
Db 64 TTGTATGAGGAATAATTTTCAGCTAGGTATTCGATATATTACAAATATGATTATGTC 123  
QY 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr 83  
|||||  
Db 124 GTAGGTGACTATGTTACGATCAATTTACGCTTAAAGGATTTTATAAAGATAGATAATAA 183  
QY 84 IleArgLys---AsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102  
|||  
Db 184 AAGCGAGAATAAGTAACGATTTTCAAGTATATACAAGATATATACGAATATTGTAAT 243  
QY 103 PheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle-----LysPhe 120  
|||||  
Db 244 TTGGTGTGCTTATTTTGTAGTCACAGCATGCTCACCANAATGAATTTATTGAAGAAATA 303  
QY 121 AspAspGlu 123  
|||||  
Db 304 GATGATAAA 312

RESULT 3

US-08-931-999-4  
; Sequence 4, Application US/08931999  
; Patent No. 6043219  
; GENERAL INFORMATION:  
; APPLICANT: Iandolo, John J.  
; APPLICANT: Crupper, Scott S.  
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hovey, Williams, Timmons & Collins  
; STREET: 2405 Grand Boulevard, Suite 400  
; CITY: Kansas City  
; STATE: Missouri  
; COUNTRY: U.S.A.  
; ZIP: 64108  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/931,999  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/710,561  
; FILING DATE: 19-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Collins, John M.  
; REGISTRATION NUMBER: 26,262  
; REFERENCE/DOCKET NUMBER: 25043-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 816/474-9050  
; TELEFAX: 816/474-9057  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6755 base pairs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 10:38:22 : Search time 49 seconds

(without alignments)

1201.672 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 1009

Sequence: 1 MRKEVTPMLNKNYPGQF.....EEQDKEMTSKOHLLFVRKN 192

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09868352/runat\_14022003\_153516\_4244/app\_query.fasta\_1.391  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORMEXT=HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868352.ecgn\_1\_1\_25-runat\_14022003\_153516\_4244 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -FGAPOP=6 -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued\_Patents\_NA.\*
- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfilesl.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	438	43.4	5558	4	US-08-961-527-103
2	254.5	25.2	399	4	US-09-134-001C-2057
3	95	9.4	6755	3	US-08-931-999-4
4	94	9.3	2790	2	US-08-718-661-1
5	87.5	8.7	333	4	US-09-018-584A-27
6	87	8.6	2017	1	US-07-667-276A-3
7	87	8.6	2237	4	US-08-914-999-7
8	86	8.5	291	1	US-07-922-723A-7
9	86	8.5	291	1	US-07-799-828C-7
10	86	8.5	291	1	US-08-074-275-7
11	86	8.5	291	1	US-08-480-366-7
12	86	8.5	291	2	US-07-952-277A-7

13	86	8.5	1818	4	US-09-108-010B-1	Sequence 1, Appl
14	85.5	8.5	443	4	US-09-397-787-194	Sequence 194, App
15	83.5	8.3	454	2	US-08-623-906A-6	Sequence 6, Appl
16	83.5	8.3	44453	4	US-09-146-053-5	Sequence 5, Appl
17	83	8.2	913	1	US-08-109-391A-1	Sequence 1, Appl
18	83	8.2	913	1	US-08-459-019A-1	Sequence 1, Appl
19	83	8.2	913	2	US-08-460-428A-1	Sequence 1, Appl
20	83	8.2	913	3	US-08-458-860A-1	Sequence 1, Appl
21	83	8.2	2354	4	US-09-810-671-1	Sequence 1, Appl
22	82.5	8.2	711	1	US-08-307-499-49	Sequence 49, Appl
23	82.5	8.2	711	4	US-09-299-268-49	Sequence 49, Appl
24	82.5	8.2	3143	1	US-08-485-621-1	Sequence 1, Appl
25	82.5	8.2	3143	2	US-08-973-831-1	Sequence 1, Appl
26	82.5	8.2	3143	5	PCT-US96-09530A-1	Sequence 1, Appl
27	82.5	8.2	5285	4	US-09-402-328-1	Sequence 1, Appl
C 27	82.5	8.2	14176	1	US-08-307-499-1	Sequence 1, Appl
C 28	82.5	8.2	14176	1	US-08-307-499-14	Sequence 14, Appl
29	82.5	8.2	14176	4	US-09-299-268-1	Sequence 1, Appl
C 30	82.5	8.2	14176	4	US-09-299-268-14	Sequence 14, Appl
31	82.5	8.2	3066	4	US-09-086-912-1	Sequence 1, Appl
32	82	8.1	3066	4	US-09-203-453-1	Sequence 1, Appl
33	82	8.1	3066	4	US-08-961-527-195	Sequence 195, App
C 34	81.5	8.1	6593	4	US-08-759-444-2	Sequence 2, Appl
35	81.5	8.1	8878	1	US-08-680-897-1	Sequence 1, Appl
C 36	81.5	8.1	9880	3	US-08-961-527-167	Sequence 167, App
37	81	8.0	9711	4	US-08-691-563C-46	Sequence 46, Appl
C 38	80.5	8.0	1859	3	US-09-470-512A-3	Sequence 3, Appl
39	80.5	8.0	41708	4	US-09-134-001C-2695	Sequence 2695, Ap
40	80	7.9	789	4	US-09-108-010B-2	Sequence 2, Appl
41	80	7.9	1920	4	US-08-306-691B-23	Sequence 23, Appl
42	80	7.9	2301	1	PCT-US93-06251-78	Sequence 78, Appl
43	80	7.9	2433	4	US-09-540-824-24	Sequence 24, Appl
44	80	7.9	2433	4	US-08-560-398-11	Sequence 11, Appl
45	80	7.9	2939	2		

ALIGNMENTS

RESULT 1

US-08-961-527-103/c

; Sequence 103, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 103:

; SEQUENCE CHARACTERISTICS:

**THIS PAGE BLANK (USPTO)**

```

    |||    |||    :::    |||||:::|||||
Db 3092 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 3033
Qy 163 ThrLysArgGlnGlnGluArgHisIleLysGluGlnGlnAspLysGluMetThrSerAla 182
Db 3032 AGAAAGAAAGAAAGGAAAGA-----AAGAAAGAAAGAAAGAAAGAAAGAAAGGA 2979
Qy 183 LysGln 184
Db 2978 AAGAAA 2973
    |||:::
```

Search completed: February 25, 2003, 13:04:24  
Job time : 95 secs

```
Db 1192 AGTGGTAGTAAAGAAAGAAATCATAGAGATTCGCCAAGTAGACAATTAATTGTTTCAAAAT 1251
      ::::: ||| ::::: |||
Qy 176 -----AspLysGluMetThrSerAlaLys 183
      ::::: ||| ::::: |||
Db 1252 GATGGTAATACATTAATACAAATAAAGAGAGATCAAAATCAAAA 1296
      ::::: ||| ::::: |||

RESULT 39
US-09-994-485-7
; Sequence 7, Application US/09994485
; Patent No. US20020142429A1
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey G.
;            Pavur, William N.
;            Hait, William N.
; TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)
; AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
;            Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/994,485
; FILING DATE: 27-NO. US20020142429A1-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/914,999
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 601-1-078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2237 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Dictyostellium discoideum
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-994-485-7

Alignment Scores:
Pred. No.: 0.423 Length: 2237
Score: 87.00 Matches: 45
Percent Similarity: 34.42% Conservative: 29
Best Local Similarity: 20.93% Mismatches: 69
Query Match: 8.62% Indels: 72
DB: 10 Gaps: 11

US-09-868-352-23 (1-192) x US-09-994-485-7 (1-2237)

Qy 12 TyrAsnLysTyrProGly-----ProGln---PheIle 21
      |||||
Db 739 TATAATAATAATAGTGATTCGTTAGTAATGATGAAGAAATAACACCAATCATTCCTCT 798
      |||||

Qy 22 HisPheGluAsnIleValLysSerAspPheIleGluPheGlnLeuValIleAsnGluLys 41
```

```
Db 799 CATTTTC-----ACATATGAACATTCAAATCATCAATTATTGATT----- 837
      |||||
Qy 42 SerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeu---LysTyr 60
      |||||
Db 838 -----ATAGATATTCAAGGTGTTGGTGATCACTATACAGACCCACAATTCATACCTAT 891
      |||||
Qy 61 Asp-----PheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPhe 77
      |||
Db 892 GATGCTGTTGGTTTGGTATTGGTAATTGGGTCAA-----AAAGGTTTT 936
      |||
Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
      |||
Db 937 GAAAGAGTTTTAGATACATCATAA----- 960
      |||
Qy 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArg--- 116
      |||||
Db 961 -----TGTAATGCAATTTGCCAATATTAAATTTTACAATCAATTAATCCAAAATCT 1011
      |||||
Qy 117 -----AspIleLysPheAspGluArgPro 125
      |||||
Db 1012 GAAAAAGTGAATTGGTACTGTACCAAGACCAGATTTAAATTTCCCTGATACATCTGAA 1071
      |||||
Qy 126 HisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsn 145
      |||||
Db 1072 AGAGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 1131
      |||||
Qy 146 ArgSerGlnSerAsnAlaHisPheThrSerLysLys-----Arg 160
      |||||
Db 1132 AATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATTC 1191
      |||||
Qy 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGluGln----- 175
      |||||
Db 1192 AGTGTAGTAAAGAAAGAAATGATAGAGATTCGCAAGTAGACAAATTAATTGTTTCAAA 1251
      |||||
Qy 176 -----AspLysGluMetThrSerAlaLys 183
      |||||
Db 1252 GATGTAATACATTAATAATAATAAGAGAGATCAAAATCAAAA 1296
      |||||

RESULT 40
US-09-764-869-1279/c
; Sequence 1279, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1279
; LENGTH: 8317
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1279

Alignment Scores:
Pred. No.: 2.73 Length: 8317
Score: 87.00 Matches: 19
Percent Similarity: 50.00% Conservative: 12
Best Local Similarity: 30.65% Mismatches: 29
Query Match: 8.62% Indels: 2
DB: 10 Gaps: 1

US-09-868-352-23 (1-192) x US-09-764-869-1279 (1-8317)

Qy 123 GluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThr 142
      |||||
Db 3152 GAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 3093
      |||||

Qy 143 ArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLysLysArgLysAsp 162
```

Qy	12	TyrAsnLysTyrProGly		----	ProGln	----	Phelle	21
Db	739	TATAATAATAATAGTGATTCGGTTAGTAATGATGAAGAAATACACCAACATCATCTCT		----	ProGln	----	Phelle	21
Qy	22	HispheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLys		----	ProGln	----	Phelle	21
Db	799	CATTTC		----	ProGln	----	Phelle	21
Qy	42	SerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeu		----	ProGln	----	Phelle	21
Db	838	-----ATAGATATTCAAGGTGGTGATCCTATACAGACCCACAAATTCATACCTAT		----	ProGln	----	Phelle	21
Qy	61	Asp		----	ProGln	----	Phelle	21
Db	892	GATGGTGTGGTTTGGTATTGGTAATTTGGGTCAA		----	ProGln	----	Phelle	21
Qy	78	TyrlsAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle		----	ProGln	----	Phelle	21
Db	937	GAAAAGTTTTAGATACTCATAA		----	ProGln	----	Phelle	21
Qy	98	LysGlnTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArg		----	ProGln	----	Phelle	21
Db	961	-----TGTAATGCATTTGCCAATATTAAATTTACAATCAATTAATCCAAATCT		----	ProGln	----	Phelle	21
Qy	117	-----AspIleLysPheAspGluArgPro		----	ProGln	----	Phelle	21
Db	1012	GAAAAAGTGATTCGTACTGTACCAAGACCAGATTAATTTTCCTGCATACATCTGAA		----	ProGln	----	Phelle	21
Qy	126	HisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsn		----	ProGln	----	Phelle	21
Db	1072	AGAGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT		----	ProGln	----	Phelle	21
Qy	146	ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLys		----	ProGln	----	Phelle	21
Db	1132	AATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT		----	ProGln	----	Phelle	21
Qy	161	LysAspThrLysArgArgGlnGluArgHisIleLysGluGln		----	ProGln	----	Phelle	21











;> LENGTH: 7521  
;> TYPE: DNA  
;> ORGANISM: Homo sapiens  
;> FEATURE:  
;> NAME/KEY: CDS  
;> LOCATION: (1)..(7521)  
US-09-819-104A-3

Alignment Scores:  
Pred. No.: 1,16 Length: 7521  
Score: 89.50 Matches: 29  
Percent Similarity: 40.85% Conservativeness: 29  
Best Local Similarity: 20.42% Mismatches: 53  
Query Match: 8.87% Indels: 31  
DB: 9 Gaps: 5

US-09-868-352-23 (1-192) x US-09-819-104A-3 (1-7521)

QY 72 LeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSer 91  
Db 1225 ATCAACATGAACGGCTTATGGCCGACCCCATGAAGGTGTACAAAGACCCGCCAGGTCTATG 1284  
QY 92 ArgLeu-----GluAspTyrIleLysGluTyrCys 101  
Db 1285 AACATGTGGAGTCAGCAGGAGAGACCTTCCGGGAGAGTTCAATGAGCATCCCAAG 1344  
QY 102 AsnPheGly-----CysAla----- 106  
Db 1345 AACTTTGGCCTGATCGCATCTTCCTGGAGAGGAGACAGTGGCTGAGTGGCTCTCTAT 1404  
QY 107 ---TyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspGluArgPro 125  
Db 1405 TACTACTGCTACTAAGAGAAATGAGAACTATAGAGCCTG-----GTGAGACGG 1452  
QY 126 HisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsn 145  
Db 1453 AGCTATCGCGCGCGCAGAGCCAGCAGCAACACAGCAGCAGCAGCAGCAGCAGCAG 1512  
QY 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165  
Db 1513 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1569  
QY 166 ArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185  
Db 1570 AAGAAAGAGAGCGGAG 1629  
QY 186 LeuLeu 187  
Db 1630 CTCCTC 1635

RESULT 27  
US-09-819-104A-1  
;> Sequence 1, Application US/09819104A  
;> Publication No. US20030027137A1  
;> GENERAL INFORMATION:  
;> APPLICANT: Chen, J. Don  
;> TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COREPRESSOR MOLECULES  
;> FILE REFERENCE: US/09/819,104A  
;> CURRENT APPLICATION NUMBER: 2001-03-27  
;> PRIOR FILING DATE: 2001-03-27  
;> PRIOR APPLICATION NUMBER: 60/193,138  
;> PRIOR FILING DATE: 2000-03-29  
;> NUMBER OF SEQ ID NOS: 6  
;> SOFTWARE: PatentIn Ver. 2.0  
;> SEQ ID NO 1  
;> LENGTH: 8686  
;> TYPE: DNA  
;> ORGANISM: Homo sapiens  
;> FEATURE:  
;> NAME/KEY: CDS  
;> LOCATION: (157)..(7677)  
US-09-819-104A-1

Alignment Scores:  
Pred. No.: 1,43 Length: 8686  
Score: 89.50 Matches: 29  
Percent Similarity: 40.85% Conservativeness: 29  
Best Local Similarity: 20.42% Mismatches: 53  
Query Match: 8.87% Indels: 31  
DB: 9 Gaps: 5

US-09-868-352-23 (1-192) x US-09-819-104A-1 (1-8686)

QY 72 LeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSer 91  
Db 1381 ATCAACATGAACGGCTTATGGCCGACCCCATGAAGGTGTACAAAGACCCGCCAGGTCTATG 1440  
QY 92 ArgLeu-----GluAspTyrIleLysGluTyrCys 101  
Db 1441 AACATGTGGAGTCAGCAGGAGAGACCTTCCGGGAGAGTTCAATGAGCATCCCAAG 1500  
QY 102 AsnPheGly-----CysAla----- 106  
Db 1501 AACTTTGGCCTGATCGCATCTTCCTGGAGAGGAGACAGTGGCTGAGTGGCTCTCTAT 1560  
QY 107 ---TyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspGluArgPro 125  
Db 1561 TACTACTGCTACTAAGAGAAATGAGAACTATAGAGCCTG-----GTGAGACGG 1608  
QY 126 HisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsn 145  
Db 1609 AGCTATCGCGCGCGCAGAGCCAGCAGCAACACAGCAGCAGCAGCAGCAGCAGCAG 1668  
QY 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165  
Db 1669 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1725  
QY 166 ArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185  
Db 1726 AAGAAAGAGAGCGGAG 1785  
QY 186 LeuLeu 187  
Db 1786 CTCCTC 1791

RESULT 28  
US-09-960-352-5438/c  
;> Sequence 5438, Application US/09960352  
;> Patent No. US20020137139A1  
;> GENERAL INFORMATION:  
;> APPLICANT: Tao, Nengbing  
;> APPLICANT: Byatt, John C.  
;> APPLICANT: Mathialagan, Nagappan  
;> TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
;> TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
;> FILE REFERENCE: 16511.006/37-21(10298)C  
;> CURRENT APPLICATION NUMBER: US/09/960,352  
;> CURRENT FILING DATE: 2001-09-24  
;> NUMBER OF SEQ ID NOS: 15112  
;> SEQ ID NO 5438  
;> LENGTH: 203  
;> TYPE: DNA  
;> ORGANISM: Bos taurus  
;> OTHER INFORMATION: Clone ID: 23-LIB34-075-01-EI-F11  
US-09-960-352-5438

Alignment Scores:  
Pred. No.: 0,00791 Length: 203  
Score: 89.00 Matches: 13  
Percent Similarity: 62.96% Conservativeness: 21  
Best Local Similarity: 24.07% Mismatches: 20  
Query Match: 8.82% Indels: 0  
DB: 10 Gaps: 0

Alignment Scores:	
Pred. No.:	0.091
Score:	90.00
Percent Similarity:	37.88
Best Local Similarity:	23.74
Query Match:	8.92%
DB:	9

Percent Similarity:	37.88%
Best Local Similarity:	23.74%
Query Match:	8.92%
DB:	9

Db	868	GGACGGAGATCAGTAGACA	:::
Qy	64	ValGlyAspTrpGlyAsnG	:::
Db	913	CTTGTGGACTGGGCTCGTC	
Qy	73	---AraLeuArgGlypHet	

D6 973 CCGAGACTTGAAGGTCAGT

Qy	85	-----ArgL	
Db	1033	CAATGCTTAAGCCATCGAC	
Qy	94	GUAsPtyrIleLysGLUty	
Db	1093	AATGAT---CTTAAAGACT	
Qy	114	AsnProGAspIleLysp	
Db	1150	ACACTCCGGACATTAAG	
Qy	133	SerLysSerGlnSerSerL	
		:	
Db	1210	CATCATCATCATCAAC	

Qy	153	AlaHisPheThrSerLeuLysL
Db	1270	ACCAAGTCCTCCATCTCCACA
RESULT	26-	
US-09-819-104A-3		
:	Sequence 3, Application US/	
:	Publication No. US200300271	
:	GENERAL INFORMATION:	
:	APPLICANT: Chen, J. Don	
:	TITLE OF INVENTION: NOVEL	
:	TITLE OF INVENTION: AND U	
:	FILE REFERENCE: UMG-030	
:	CURRENT APPLICATION NUMBER	
:	PRIOR FILING DATE: 2001	
:	CURRENT APPLICATION NUMBER:	
:	PRIOR FILING DATE: 2000-03	
:	NUMBER OF SEQ ID NOS: 6	
:	SOFTWARE: PatentIn Ver. 2.	
:	SEQ. ID. NO. 3	

ArgLysAspThrLysArgArgGlnGluArgHis 170  
||| ::| ||| |||  
GCCAAGTCGCCGAGGAATTCAACGGAATCAC 1323

819104A  
A1

CLEAR RECEPTOR COREPRESSOR MOLECULES  
S THEREFOR

US/09/819,104A  
3-27  
/193,138  
9



```
Db 654 ATTAGGCTTATTGAAGATACCAAGCCACAGCCATAGCGATCTTACTCTGGAAGCAGA 713
QY 126 -----HisLysArgArgLysSerArgSerLysSer---GlnSerSerLysSerGln 141
Db 714 TCCAGGTCTCGATCTAGAACACGGTCACGAAGTAGGAGTCGACGAGACGCCGCGTAGA 773
QY 142 ThrArgAsn---AsnArgSerGlnSerAsnAlaAlaHisPheThrSerLysLysArg 160
Db 774 TCTCGAAGTATCTCAAAAGTCGCTCCGTTCCAGGTCCGGAGCAAAAGTCGATCAGT 833
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThr 180
Db 834 TCTCGATCAAAAGCAGCAAAATCTAGATCAAAAGCAAAATCTAAGCCCAAGTCTGATCG 893
QY 181 SerAlaLysGlnHis 185
Db 894 GGTCTCCCATTCACAT 908

RESULT 21
US-09-728-952-78
; Sequence 78, Application US/09728952
; Patent No. US20020111302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/728,952
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: pt_FL-genes Version 2.0
; SEQ ID NO 78
; LENGTH: 5273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)..(4832)
US-09-728-952-78

Alignment Scores:
Pred. No.: 0-527 Length: 5273
Score: 90.50 Matches: 46
Percent Similarity: 34.90% Conservative: 21
Best Local Similarity: 23.96% Mismatches: 62
Query Match: 8.97% Indels: 63
DB: 10 Gaps: 7

US-09-868-352-23 (1-192) x US-09-728-952-78 (1-5273)
QY 24 GluAsnIleValLysSerAspAlleGluPheGlnLeuValIleAsnGluLysSerAla 43
Db 4278 AAAAACAATTGTTCAA---GATGACAAAGAGGTGATCTTGAGCTCAGAGGAGGAGTTTC 4334
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIle 63
Db 4335 TTT-----GTCCAA 4343
QY 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr 83
Db 4344 GTCCATGATGTTTCTCCAGAGCAACCTCGA-----ACA 4376
QY 84 IleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGlu---TyrCysAsn 102
Db 941 GTCCAA 4941
```

```
Db 4377 GTCATAAAGCACCCCGCGTCAGCACTGCACAGGATGTCATTTCAGCAGACCTTTATGCAA 4436
QY 103 PheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspAsp 122
Db 4437 GCCAAATATTCCTACAGCATCCTGACGACACCCCAANTCCAAGCGACTATGCTTTGGAA 4496
QY 123 Glu-----ArgProHisLysArgArgLysSerArgSerLysSerGln----- 136
Db 4497 GAGGTGGTGAAGACACTACCAACAAGAGACTACCACACCAAGTCTCTCAGCGGTC 4556
QY 136 ----- 136
Db 4557 CTTCTGGATCAGCAGTGTGTTCAGGCCAAAGTGAAGAGTGCAGGAAATTC 4616
QY 137 -----SerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHis 154
Db 4617 ATCCTTAAGCTAAAGGAGCAGGTGCAGGCATCTCGAGAGATATAAAGAAAGCATTTCT 4676
QY 155 PheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHisIleLysGluGlu 174
Db 4677 TTGCAAGTGAAGTCAAGAGCTCACCAG-----TCAACTAAA 4715
QY 175 GlnAspLysGluMetThrSerAlaLysGlnHisLeu 186
Db 4716 CAGCCCGAGGACTTACATCACCTTCTCAGCTCTTG 4751

RESULT 22
US-09-822-635-3
; Sequence 3, Application US/09822635
; Patent No. US20010039331A1
; GENERAL INFORMATION:
; APPLICANT: Hunter, John J.
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 16836, A NOVEL HUMAN PHOSPHOLIPASE C AND
; FILE REFERENCE: USES THEREOF
; FILE REFERENCE: 10448-Q35001
; CURRENT APPLICATION NUMBER: US/09/822,635
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 60/193,921
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-635-3

Alignment Scores:
Pred. No.: 0-55 Length: 5430
Score: 90.50 Matches: 46
Percent Similarity: 34.90% Conservative: 21
Best Local Similarity: 23.96% Mismatches: 62
Query Match: 8.97% Indels: 63
DB: 10 Gaps: 7

US-09-868-352-23 (1-192) x US-09-822-635-3 (1-5430)
QY 24 GluAsnIleValLysSerAspAlleGluPheGlnLeuValIleAsnGluLysSerAla 43
Db 4876 AAAAACAATTGTTCAA---GATGACAAAGAGGTGATCTTGAGCTCAGAGGAGGAGTTTC 4932
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIle 63
Db 4933 TTT-----GTCCAA 4941
QY 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr 83
Db 4942 GTCCATGATGTTTCTCCAGAGCAACCTCGA-----ACA 4974
QY 84 IleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGlu---TyrCysAsn 102
Db 4975 GTCATCAAGCACCCCGCGTCAGCACTGCACAGGATGTCATTTCAGCAGACCTTTATGCAA 5034
```



Qy	174	GluGlnAspLysGluMetThrSerAlaLys	183
		:::	::::::::::
Db	212	AGGGGTAAAGGTCAGGAGCAGCGTCAAAA	241
RESULT 20			
US-10-002-600-58			
: Sequence 58, Application US/10002600			
: Patent No. US20020137077A1			
: GENERAL INFORMATION:			
: APPLICANT: Hopkins, Christopher M.			
: APPLICANT: Peterson, David P.			
: APPLICANT: Cocks, Benjamin G.			
: APPLICANT: Hawkins, Phillip R.			
: TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS			
: FILE REFERENCE: PA-0042 US			
: CURRENT APPLICATION NUMBER: US/10/002,600			
: CURRENT FILING DATE: 2001-10-25			
: PRIOR APPLICATION NUMBER: 60/243,521			
: PRIOR FILING DATE: 2000-10-25			
: NUMBER OF SEQ ID NOS: 116			
: SOFTWARE: PERL Program			
: SEQ ID NO 58			
: LENGTH: 3697			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
: FEATURE:			
: NAME/KEY: misc_feature			
: OTHER INFORMATION: Template ID: 039976.16			
: NAME/KEY: unsure			
: LOCATION: 1444, 1446, 1448-1449, 1454, 1456			
: OTHER INFORMATION: a, t, c, g, or other			
US-10-002-600-58			
Alignment Scores:			
Pred. No.:	0.276	Length:	3697
Score:	91.00	Matches:	52
Percent Similarity:	38.22%	Conservative:	34
Best Local Similarity:	23.11%	Mismatches:	83
Query Match:	9.02%	Indels:	56
DB:	12	Gaps:	9
US-09-868-352-23 (1-192) x US-10-002-600-58 (1-3697)			
Qy	17	GlyProGlnPheIleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeu	36
		:::	::::::::::
Db	234	GGGTACGGCTTCGTTCGAGTTCGAGGACTCCCGCAGCGCAGCAGCGCGTTTACGAGCTG	293
Qy	37	ValIleAsnGluLysSerAlaPheAspValThrVal	53
		:::	::::::::::
Db	294	AACGGCAAGGAGCTCTCGCGCAGCGCTGATCGTAGAGCACGCCCGGGCGCGCGTCGC	353
Qy	54	SerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArg	73
		:::	::::::::::
Db	354	GATCGCAGCGCTACAGCTACGGAAGCGGAGNGTGGAGGTGATACACAGTCGGAGA	413
Qy	74	LeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIle	90
		:::	::::::::::
Db	414	ACATCTGSCACAGACAAATACGGACCACCTGTTCTGTACAGAAATACAGGCTTATTGTAGAA	473
Qy	91	-----SerArg-----LeuGluAspTyrIle	97
		:::	::::::::::
Db	474	AATCTTTCTAGTCGGTGCAGTTGGCAAGATTTAAAGGATTTTATGCCAAGCAGGTGAA	533
Qy	98	-----LysGluTyrCysAsnPheGlyCysAlaTyrPhe	108
		:::	::::::::::
Db	534	GTACCTATGCGGATGCCACAGAGCAACAAATGAGGCTGTAATGTAGTTTCGCTCC	593
Qy	109	-----ValLeuGlu-----AsnProAsnProArgAsp	117
		:::	::::::::::
Db	594	TACTCTGACATGAAGCGTGTCTTGGACAAACTGGATGGCAGAAATAATATGCAGAAAT	653
Qy	118	IleLysPheAspAspGluArgPro	125
		:::	::::::::::

```

> PCT/US01/00663
> PCT FILING DATE: 2001-01-30
> PCT APPLICATION NUMBER: PCT/US01/00662
> PCT FILING DATE: 2001-01-30
> PCT APPLICATION NUMBER: PCT/US01/00661
> PCT FILING DATE: 2001-01-30
> PCT APPLICATION NUMBER: PCT/US01/00670
> PCT FILING DATE: 2001-01-30
> PCT APPLICATION NUMBER: US 60/234,687
> PCT FILING DATE: 2000-09-21
> PCT APPLICATION NUMBER: US 09/608,408
> PCT FILING DATE: 2000-06-30
> PCT APPLICATION NUMBER: US 09/774,203
> PCT FILING DATE: 2001-01-29
> NUMBER OF SEQ ID NOS: 49117
> SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
> SEQ ID NO 21362
> LENGTH: 1134
> TYPE: DNA
> ORGANISM: Homo sapiens
> FEATURE:
> OTHER INFORMATION: MAP TO AC006059.3
> OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
> OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
> OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
> OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
> OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
> OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
> OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
> OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
> OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
> OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
> OTHER INFORMATION: NT HIT: AF184110.1, EVALUAE 0.00e+00
> OTHER INFORMATION: EST HUMAN HIT: A1801430.1, EVALUAE 1.00e-127
> OTHER INFORMATION: SWISSPROT HIT: P30414, EVALUAE 1.00e-127
US-09-864-761-21362

```

Alignment Scores:			
Pred. No.:	0.0447	Length:	1134
Score:	91.50	Matches:	32
Percent Similarity:	40.35%	Conservative:	14
Best Local Similarity:	28.07%	Mismatches:	41
Query Match:	9.07%	Indels:	27
DB:	10	Gaps:	4
US-09-868-352-23 (1-192) x US-09-864-761-21362 (1-1134)			
Qy	78	TyrlYsAsp-----AlaSerThrIleArgLysAsnSerArgIleSerArgL	
Db	594	TACAGTCATAGAGTTCAGAAAGCTCACCAGAGTCAAGGACAGATCTTCTCAGAA	
Qy	95	AspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnPr	
		:	-----
Db	634	TCTTATTCCAGATCATACAAGA-	-----
Qy	115	ProArgSpIleLysPheAspAspGluArgProHisLysArgArgLysSerArgSS	
Db	610	-----TCACGTAGTCTAGCTAGTTCACATTCACAGGT	
Qy	135	SerGlnSerSerLysSerGlnThrArgAsn-----AsnArgSerGlnSerAr	
		:	
Db	574	TCTCCATCATCTAGATCTCATTCACGAAATAATACAGTCATCATTCACAGTGA	
Qy	152	AsnAlaHisPheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgH	
		:	
Db	514	TCATCTTCATATCTTCTTATTAGC---AGTCATCATGGAAGCGAGCTAAGAGGA	
Qy	172	LysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis	.185
		:	
Db	457	AGATCCAGTCGGAAAAAATAATAGCGTTTCACATAAAAGCAT	416
RESULT 18			
US-09-864-761-4620/c			

```

: Sequence 4620, Application US/09864761
: Patent NO. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Rank, David K.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIDE SEQUENCES
: FILE REFERENCE: Aesomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
: SEQ ID NO 4620
: LENGTH: 1960
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC006059.3
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
: US-09-864-761-4620

Alignment Scores:
Pred. No.: 0.0972 Length: 1960
Score: 91.50 Matches: 32
Percent Similarity: 40.35% Conservative: 14
Best Local Similarity: 28.07% Mismatches: 41

```

```
; Sequence 4310, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 4310
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(584)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3028-014-Q1-B1-F10
US-09-878-574-4310

Alignment Scores:
Pred. No.: 0.0131 Length: 584
Score: 92.50 Matches: 20
Percent Similarity: 55.74% Conservative: 14
Best Local Similarity: 32.79% Mismatches: 18
Query Match: 9.17% Indels: 9
DB: 10 Gaps: 2

US-09-868-352-23 (1-192) x US-09-878-574-4310 (1-584)
QY 134 LysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAla 153
   ||| |||::: ::::: |||::: ::::: ||| |||:::
Db 583 AAAAAACACAGACACACAAAAACCAACACACACACACAAAAACGAGAAACA 524

QY 154 HisPheThrSerLysLysArgLysAsp-----ThrLysArgArg 166
   ||| ||| ||| ||| |||::: |||::: |||::: |||:::
Db 523 CACACACACCAAAAAACAAAAACAAAAACACACACAAAAACAAAAACAAAAA 464

QY 167 GluGluArgHisLysGlu-----GluGlnAspLysGluMetThrSerAlaLysGln 184
   |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 463 CAAAAACACACAAAAACAAAAACCAACAAAAACAAAAACACACAAAAACAA 404

QY 185 His 185
   |||
Db 403 CAC 401

RESULT 16
US-10-001-843-45
; Sequence 45, Application US/10001843
; Patent No. US20020132255A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0267
; CURRENT APPLICATION NUMBER: US/10/001,843
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,992
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 1267
```

```
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (358)..(358)
; OTHER INFORMATION: a, c, g or t
; NAME/KEY: misc.feature
; LOCATION: (478)..(478)
; OTHER INFORMATION: a, c, g or t
US-10-001-843-45

Alignment Scores:
Pred. No.: 0.0393 Length: 1267
Score: 92.50 Matches: 20
Percent Similarity: 59.42% Conservative: 21
Best Local Similarity: 28.99% Mismatches: 25
Query Match: 9.17% Indels: 3
DB: 12 Gaps: 1

US-09-868-352-23 (1-192) x US-10-001-843-45 (1-1267)
QY 119 LysPheAspGluArgProHisLysArgLysSer-----ArgSerLysSer 135
   ||| ||| |||::: |||::: |||::: |||:::
Db 802 AAAAAACAGACAAAGAAAAACAAAAACAAAAACAAAAACAAAAACAGAA 861

QY 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
   ||| ||| |||::: |||::: |||::: |||:::
Db 862 CAATAAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAGAGA 921

QY 156 ThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHisLysGluGln 175
   ||| ||| |||::: |||::: |||::: |||::: |||::: |||:::
Db 922 GAGAGAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAGAAAAA 981

QY 176 AspLysGluMetThrSerAlaLysGln 184
   ::|||::: |||::: |||:::
Db 982 AACAAAAACAAAAACAAAAACAAAAACAAAAA 1008

RESULT 17
US-09-864-761-21362/c
; Sequence 21362, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
```





```
; Sequence 1933, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; PRIOR FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1933
; LENGTH: 712
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-550-1933

Alignment Scores:
Pred. No.: 0.00271 Length: 712
Score: 99.00 Matches: 29
Percent Similarity: 46.00% Conservative: 17
Best Local Similarity: 29.00% Mismatches: 22
Query Match: 9.81% Indels: 32
DB: 10 Gaps: 4

US-09-868-352-23 (1-192) x US-09-867-550-1933 (1-712)

Qy 115 ProArgAspIleLysPheAspGluArgProHis----- 126
Db 271 CCTCGTTCACATCTTATGATAGAGACGCGCATCGATCAAGCAGTAGCTCTTCTTAT 330
Qy 127 -----LysArgLysSerArgSerLysSerGlnSer----- 137
Db 331 GCCTCCAGAGAAACGAAGTCGAAGTCGTTCAAGGGGTGAGGAAATCCTATAGATT 390
Qy 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 153
Db 391 CAGAGGCTAGGTCAAAAAGCAGAAACAAAGAGGTCCAGGTCAAGACCTCGTCTCGTCT 450
Qy 154 HisPheThrSerLysLysArg-----LysAspThr 163
Db 451 CATAGCTGATGAGTGAAGGTCCAGTCACAGAAAGCGGTAGTCGTCTCGGGATAGA 510
Qy 164 LysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSerAlaLys 183
Db 511 GAACGACGTAGGGCAGAGAT---AAAGAGAAAGAGAAAGGAGAGGATAAAGGGAG 567

RESULT 8
US-09-925-301-557
; Sequence 557, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 557
; LENGTH: 3484
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
US-09-925-301-557

Alignment Scores:
Pred. No.: 0.0259 Length: 3484
Score: 99.00 Matches: 24
Percent Similarity: 58.57% Conservative: 17
Best Local Similarity: 34.29% Mismatches: 21
Query Match: 9.81% Indels: 8
DB: 10 Gaps: 2

US-09-868-352-23 (1-192) x US-09-925-301-557 (1-3484)

Qy 122 AspGluArgProHisLysArgLysSerArgSerLysSerGlnSer----- 137
Db 1101 GATAGAAAGAGAAAAAGAGGCATTCAAGATCAAGTACGTTCTAGGAGGAGGAGG 1160
Qy 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
Db 1161 ACTCCCTCATCTTCTAGACACAGCGGGTCAAGAAGCAGATCGAGACGGCGGTCACTTCT 1220
Qy 156 ThrSerLysLysArgLysAspThrLys-----ArgArgGlnGluArgHisIleLysGlu 173
Db 1221 AAGTCTAGGATCGGCGCAGATCCAAAAGCCCAAGCGGCGAAGATCTCATTTCCAGAGAA 1280
Qy 174 GluGlnAspLysGluMetThrSerAlaLys 183
Db 1281 AGAGGTAGAGGTCAAGGAGGCACATCAAAA 1310

RESULT 9
US-09-834-975-822
; Sequence 822, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew
; APPLICANT: Van Huffel, Christophe
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-016B
; CURRENT APPLICATION NUMBER: US/09/834,975
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,538
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 822
; LENGTH: 2025
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2025)
; OTHER INFORMATION: n = A,T,C or G

US-09-834-975-822

Alignment Scores:
Pred. No.: 0.0184 Length: 2025
Score: 97.50 Matches: 19
Percent Similarity: 52.86% Conservative: 18
Best Local Similarity: 27.14% Mismatches: 32
Query Match: 9.66% Indels: 1
DB: 10 Gaps: 1

US-09-868-352-23 (1-192) x US-09-834-975-822 (1-2025)

Qy 116 ArgAspIleLysPheAspGluArgProHisLysArgArgProHisLysSerArgSerLysSer 135
Db 1587 CAGAACGCCCAACGACGACGAGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1646
Qy 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
Db 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
```

```
Db 1 ATGGTGACGATTTTCAGCATGTAATGAAGTCTCGAAACAGCGCANTGTTGGAAT 60
Qy 46 ValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGly 65
Db 61 GAGGAGCATTCAAAGCGGTACAGCGTACTGAACAAATATGATACATCGTCGCG 120
Qy 66 AspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp-----AlaSer 82
Db 121 GATTTGGGCTACAAACCACTGCGGTTCGCGCGCTTTTTCGATGACAAACAAAAATCA 180
Qy 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102
Db 181 ACATTGCATAAA-----AAATAGACGCTTACCCGACTATCTTTACGAATACTGCAAT 234
Qy 103 PheGlyCysAlaTyrPheValLeu 110
Db 235 TTTGGTTCGCTTATTTGTCCTG 258

RESULT 5
US-09-974-300-3145
; Sequence 3145, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3145
; LENGTH: 273
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-3145

Alignment Scores:
Pred. No.: 3,48e-19 Length: 273
Score: 222.50 Matches: 42
Percent Similarity: 74.68% Conservative: 17
Best Local Similarity: 53.16% Mismatches: 19
Query Match: 22.05% Indels: 1
DB: 10 Gaps: 1

US-09-868-352-23 (1-192) x US-09-974-300-3145 (1-273)
Qy 34 PheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPhe 53
Db 26 TTTGAAGTCGTCAAGATGTGAAGAAAGCGTTTAAACGAAGACGCTTTAAAGCCAGATAC 85
Qy 54 SerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArg 73
Db 86 TCCGACATTTAAATAATACGACTACATCGTCGGGATTGGGGATACACCGAGCTGAGS 145
Qy 74 LeuArgGlyPheTyrLysAspAlaSer---ThrIleArgLysAsnSerArgIleSerArg 92
Db 146 CTGAAGCGCTTTTTCGATGATGCAATTTCCGATGTGCTTCTTTGTTTAAAA 262

Qy 93 LeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111
Db 206 CTTGATGAATATATTACGAGTACTGCAATTTCCGATGTGCTTCTTTGTTTAAAA 262

RESULT 6
US-10-044-090-268
; Sequence 268, Application US/10044090
; Patent No. US20020137081A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 268
; LENGTH: 1534
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 988704.26
US-10-044-090-268

Alignment Scores:
Pred. No.: 0.007 Length: 1534
Score: 99.50 Matches: 43
Percent Similarity: 41.23% Conservative: 44
Best Local Similarity: 20.38% Mismatches: 73
Query Match: 9.86% Indels: 51
DB: 12 Gaps: 8

US-09-868-352-23 (1-192) x US-10-044-090-268 (1-1534)
Qy 20 PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsn 39
Db 293 TTTGTGGAATTTGAGGATCCAGGGATGCAGATGATGCTGTGTATGAGCTTGAGAAA 352
Qy 40 GluLysSerAlaPheAspValThrVal----- 48
Db 353 GAACCTGTGTAGTCAAAAGGTTTACTATTGAAACATGCTAGGCTCGGTACAGGATGGAAGA 412
Qy 49 -----PheGlyGlnArgPheSerGluIleLeuLysTyrAsp----- 61
Db 413 GGTAGAGGAGGATACTCTGACCGTTTGTAGTAGTCGAGACCTCGAAATGATAGAGAAAT 472
Qy 62 -----PheIleValGlyAsp----- 66
Db 473 GCTCCACCTGTAAAGACAGAAAATCGTCTTATAGTTGAGAAATTTATCTCAAGAGTCAGC 532
Qy 67 TrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArg--- 85
Db 533 TGGCAGGAT-----CTCAAAGATTTCATGAGACAAGCTGGGGAAGTAACGTTT 580
Qy 86 LysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCys 105
Db 581 GCGGATGCACACGACCTAAATTTAAATGAAGGGTGTGTTGAGTTGCTTATGTTGAC 640
Qy 106 AlaTyrPheValLeuGlu-----AsnProAsnProArgAspIleLys--- 119
Db 641 TTAAGAATCTCTATTGAAAACATCTTCTGGAAAGCAATAAATCGGAGAAAAATAAATTA 700
Qy 120 PheAspAspGluArgProHisLysArgLysSerArgSerLysSerGlnSer----- 137
Db 701 ATTGAAGGCACGAAGGACAGTGTGTCAGAAAGAGGAGTCTCGATCCCGACAGAGT 760
Qy 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsn 152
Db 761 TCCTCTAGGTCCTGATAGCCGATCCCGTTCCCGTAGTCGCAAAATCTTTACAGCCCGTCAAGA 820
Qy 153 AlaHisPheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHisIleLys 172
Db 821 ACAGAGGACGAGGACCGGACCGACCAAGTCCCGTCTCTTAGTAGTCTCCCGTGCCT 880
Qy 173 GluGluGlnAspLysGluMetThrSerAlaLys 183
Db 881 GAGAAGACCCAGAAACGTTGTTCTTCAAGTAGA 913

RESULT 7
US-09-867-550-1933
```





;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 9316  
;; LENGTH: 531  
;; TYPE: DNA  
;; ORGANISM: Streptococcus pneumoniae  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)...(531)  
US-09-815-242-9316

Alignment Scores:  
Pred. No.: 2,86e-47 Length: 531  
Score: 452.50 Matches: 94  
Percent Similarity: 66.49% Conservatives: 33  
Best Local Similarity: 49.21% Mismatches: 48  
Query Match: 44.85% Indels: 16  
DB: 10 Gaps: 4

US-09-868-352-23 (1-192) x US-09-815-242-9316 (1-531)

```
QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||
Db 1 ATCGGAAAAGAAATTCACCTGAATATATACAACTATACAAAGTTCTCGTCGGAGTTC 60

QY 21 IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40
|||||
Db 61 CATTTACACGGGACAAAGTCGAAACGGAAGGATAGCTTTTCCCTGGTTGAAATATC 120

QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
|||||
Db 121 AAGGATGCCTTTCATGTGACGACTTTTATACAGCGTTTTCAGAACTATTAAACCAAGTTT 180

QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
|||||
Db 181 GATTATATCGTGGGGACCTGGAGCAACGACGTTTCGCTACGAGGTTTTCACAAAGGAT 240

QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
|||||
Db 241 GACCGACAGAGAAACTTGAAGAAATCAGTCGTTTACAGACTACCTTTTAGAGTAT 300

QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
|||||
Db 301 TGTAGTTATGTTGCTGCTTATTTGCTTAGAAAATGAAGCCCTTACAGAGTCTTAAACCAAGTTT 180

QY 161 LysAspThrLysArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThr 180
|||||
Db 448 AAGAAAATAGACGTCGTGAGAAAGACCAAGCATTCTCAGAAAAGAGCAAGGAA----- 501

QY 181 SerAlaLysGlnHis-LeuLeuPheValArg 190
|||||
Db 502 -----CAACGTCATTCTTCATTCGTCGAC 526
```

RESULT 2

US-09-815-242-9469  
; Sequence 9469, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.

;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John D.  
;; APPLICANT: Carr, Grant J.  
;; APPLICANT: Yamamoto, Robert T.  
;; APPLICANT: Xu, H. Howard  
;; TITLE OF INVENTION: Identification of Essential Genes in  
;; TITLE OF INVENTION: Prokaryotes  
;; FILE REFERENCE: ELITRA.011A  
;; CURRENT APPLICATION NUMBER: US/09/815,242  
;; CURRENT FILING DATE: 2001-03-21  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 9469  
;; LENGTH: 531  
;; TYPE: DNA  
;; ORGANISM: Streptococcus pneumoniae  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)...(531)  
US-09-815-242-9469

Alignment Scores:  
Pred. No.: 2,86e-47 Length: 531  
Score: 452.50 Matches: 94  
Percent Similarity: 66.49% Conservatives: 33  
Best Local Similarity: 49.21% Mismatches: 48  
Query Match: 44.85% Indels: 16  
DB: 10 Gaps: 4

US-09-868-352-23 (1-192) x US-09-815-242-9469 (1-531)

```
QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||
Db 1 ATCGGAAAAGAAATTCACCTGAATATATACAACTATACAAAGTTCTCGTCGGAGTTC 60

QY 21 IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40
|||||
Db 61 CATTTACACGGGACAAAGTCGAAACGGAAGGATAGCTTTTCCCTGGTTGAAATATC 120

QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
|||||
Db 121 AAGGATGCCTTTCATGTGACGACTTTTATACAGCGTTTTCAGAACTTTAAACCAAGTTT 180

QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
|||||
Db 181 GATTATATCGTGGGGACCTGGAGCAACGACGTTTCGCTACGAGGTTTTCACAAAGGAT 240

QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
|||||
Db 241 GACCGACAGAGAAACTTGAAGAAATCAGTCGTTTACAGACTACCTTTTAGAGTAT 300

QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
|||||
Db 301 TGTAGTTATGTTGCTGCTTATTTGCTTAGAAAATGAAGCCCTTACAGAGTCTTAAACCAAGTTT 180

QY 121 AspAspGluArgProHisLysArgLysSerArgLysSerArgSerLysSerLysSer 140
|||||
Db 361 GAC-----AAGAAAATGCGTAAAGACGGAAGAAACACCTTCTAGAAAAGGAGAA 414
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 11:35:13 ; Search time 67 seconds  
(without alignments)  
1609.474 Million cell updates/sec

Title: US-09-868-352-23  
Perfect score: 1009  
Sequence: 1 MRKEVTPMLNKNKYPQPF.....EEQDKEMTSKQHLFVRKN 192

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 442118 seqs, 280819700 residues  
Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09868352/runat\_14022003\_153518\_4305/app\_query.fasta\_1.391  
-DB=Published\_Applications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09868352 -CGN\_1\_1\_33 @runat\_14022003\_153518\_4305  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA :  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq : \*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq : \*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq : \*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq : \*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq : \*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq : \*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq : \*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq : \*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq : \*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq : \*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq : \*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq : \*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq : \*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	452.5	44.8	531	10	US-09-815-242-9316 Sequence 9316, Ap
2	452.5	44.8	531	10	US-09-815-242-9469 Sequence 9469, Ap
3	302.5	30.0	1495	10	US-09-070-927A-502 Sequence 502, App
4	232.5	23.0	285	10	US-09-974-300-7936 Sequence 7936, Ap

5	222.5	22.1	273	10	US-09-974-300-3145	Sequence 3145, Ap
6	99.5	9.9	1534	12	US-10-044-090-268	Sequence 268, App
7	99	9.8	712	10	US-09-867-550-1933	Sequence 1933, Ap
8	99	9.8	3484	10	US-09-925-301-557	Sequence 557, App
9	97.5	9.7	2025	10	US-09-834-975-822	Sequence 822, App
10	96.5	9.6	1428	9	US-10-001-857-79	Sequence 79, Appl
11	95.5	9.5	1343	12	US-10-001-843-70	Sequence 70, Appl
12	95.5	9.5	2212	10	US-09-925-297-142	Sequence 142, App
13	95.5	9.5	3259	12	US-10-001-843-71	Sequence 71, Appl
14	93.5	9.3	2269	9	US-09-989-920-113	Sequence 113, App
15	92.5	9.2	584	10	US-09-878-574-4310	Sequence 4310, Ap
16	92.5	9.2	1267	12	US-10-001-843-45	Sequence 45, Appl
17	91.5	9.1	1134	10	US-09-864-761-21362	Sequence 21362, A
18	91.5	9.1	1960	10	US-09-864-761-4620	Sequence 4620, Ap
19	91	9.0	425	10	US-09-960-352-10716	Sequence 10716, A
20	91	9.0	3697	12	US-10-002-600-58	Sequence 58, Appl
21	90.5	9.0	5273	10	US-09-728-952-78	Sequence 78, Appl
22	90.5	9.0	5430	10	US-09-822-635-3	Sequence 3, Appli
23	90.5	9.0	10172	10	US-09-822-635-1	Sequence 1, Appli
24	90	8.9	272	10	US-09-864-761-18684	Sequence 18684, A
25	90	8.9	1389	9	US-09-938-842A-340	Sequence 340, App
26	89.5	8.9	7521	9	US-09-819-104A-3	Sequence 3, Appli
27	89.5	8.9	8686	9	US-09-819-104A-1	Sequence 1, Appli
28	89	8.8	203	10	US-09-960-352-5438	Sequence 5438, Ap
29	89	8.8	489	10	US-09-920-300A-1624	Sequence 1624, Ap
30	89	8.8	489	12	US-10-033-528-1624	Sequence 1624, Ap
31	89	8.8	493	10	US-09-920-300A-688	Sequence 688, App
32	89	8.8	493	12	US-10-033-528-688	Sequence 688, App
33	89	8.8	609	9	US-09-796-692-4386	Sequence 4386, Ap
34	88	8.7	170834	10	US-09-835-232-7	Sequence 7, Appli
35	87.5	8.7	333	10	US-09-784-423-27	Sequence 27, Appl
36	87	8.6	575	10	US-09-864-761-20733	Sequence 20733, A
37	87	8.6	1969	10	US-09-864-761-3972	Sequence 3972, Ap
38	87	8.6	2237	9	US-09-832-292-11	Sequence 11, Appl
39	87	8.6	2237	10	US-09-994-485-7	Sequence 7, Appli
40	87	8.6	8317	10	US-09-764-869-1279	Sequence 1279, Ap
41	86.5	8.6	400	10	US-09-983-965-265	Sequence 265, App
42	86.5	8.6	832	9	US-10-001-857-97	Sequence 97, Appl
43	86.5	8.6	1717	9	US-10-001-857-99	Sequence 99, Appl
44	86.5	8.6	172637	10	US-09-805-458A-3	Sequence 3, Appli
45	86	8.5	919	10	US-09-878-574-4303	Sequence 4303, Ap

ALIGNMENTS

RESULT 1  
US-09-815-242-9316  
; Sequence 9316, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27

**THIS PAGE BLANK (USPTO)**



JOURNAL COMMENT	Unpublished (1999)		ORGANISM				
	Contact: Sandy Clifton, Ph.D. WashU Xenopus EST project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu Library constructed by A. Zorn and J. Mason (Wellcome/CRC Institute ) . DNA Sequencing by: Washington University Genome Sequencing Center						
FEATURES source	Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov High quality sequence stop: 488. Location/Qualifiers 1. .622 /organism="Silurana tropicalis" /db_xref="taxon:8364" /clone="IMAGE:4462064" /clone_lib="Wellcome CRC pCS107 tropicalis Stc10-12" /tissue_type="whole embryo, stages 10-12" /lab_host="DH10B (phage-resistant)" /note="Vector: pCS107; Site.1: NotI; Site.2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Average insert size 1.5 kb, range 0.5-4 kb. Library constructed by A. Zorn and J. Mason (Wellcome/CRC Institute)."						
BASE COUNT	107 a	152 c 108 g 255 t	REFERENCE				
	A. Zorn and J. Mason (Wellcome/CRC Institute). " "						
ORIGIN	Alignment Scores:		REFERENCE				
	Pred. No.: 0.57 Length: 622 Score: 104.50 Matches: 30 Percent Similarity: 51.33% Conservative: 27 Best Local Similarity: 27.03% Mismatches: 41 Query Match: 10.36% Indels: 13 DB: 12 Gaps: 4						
US-09-868-352-23 (1-192) x BG885312 (1-622)							
QY	80	AspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGlu 99	REFERENCE				
Db	514	GACGAGCTCTACAGAGAACCAAGAGCCCAAC-----AAGAGG 473					
QY	100	TyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLys 119	REFERENCE				
Db	472	AAACGCAACAAAGGCGACGCCATTCACAAAGTCCGCGGATCCCGAAGCT---GACATGGAA 416					
QY	120	PheAspAsp-----GluArgProHisLysArgArgLysSerArgSerLys 134	REFERENCE				
Db	415	AATACGGATCTCCGTGTCACAGAAAGGCCACATCAAGAAAGTCAAGAAATATAA 356					
QY	135	Ser---GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAla 153	REFERENCE				
Db	355	AAATATCAGGAGACTCAAAATGAGACTGTGTGGAACATAAGATCAGGAGATGAAGAGGAG 296					
QY	154	HisPheThrSerLysLysArgLysAspThrLysArgGlnGluArgHisIleLysGlu 173	REFERENCE				
Db	295	AAACCGAAAAACGAAAAAGAAAAATAAGAGAGAGGAGATGAACCCCGGGAGGAG 236					
QY	174	GluGlnAspLysGluMetThrSerAlaLysGln 184	REFERENCE				
Db	235	GAGGAGGAGGAGGAGTTAGACTCTCACACACAG 203					
RESULT 40							
LOCUS	AK007729						
DEFINITION	Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched library, clone:1810038B01:splicing factor, arginine/serine-rich 5 (SRP40, HRS), full insert sequence.						
ACCESSION	AK007729.1 GI:12841459						
VERSION	AK007729						
KEYWORDS	HTC; CAP trapper.						
SOURCE	Mus musculus (strain:C57BL/6J) 10 day old male pancreas cDNA to						
mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library							
Mus musculus							
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
1 Carninci,P. and Hayashizaki,Y.							
High-efficiency full-length cDNA cloning							
Meth. Enzymol. 303, 19-44 (1999)							
99279253							
10349636							
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.							
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes							
Genome Res. 10 (10), 1617-1630 (2000)							
20499374							
11042159							
3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.							
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer							
Genome Res. 10 (11), 1757-1771 (2000)							
20530913							
11076861							
4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yananaka,I., Saito,T., Okazaki,Y., Gojohori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaide,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staib,D., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.							
Functional annotation of a full-length mouse cDNA collection							
Nature 409 (6821), 685-690 (2001)							
21085660							
11217851							
5 (bases 1 to 1166)							
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yananaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.							
Direct Submission							
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of							





```

/clone="IMAGE:5363670"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

```

BASE COUNT 438 a 186 c 142 g 54 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.681 Length: 820  
Score: 105.00 Matches: 27  
Percent Similarity: 57.33% Conservative: 16  
Best Local Similarity: 36.00% Mismatches: 27  
Query Match: 10.41% Indels: 5  
DB: 13 Gaps: 2

US-09-868-352-23 (1-192) x BI734757 (1-820)

Qy 113 ProAsnProArgAspIleLysPheAspGluArgProHisLysArgLysSerArg 132  
||| |||: : : : : ||| ||| |||: : :  
Db 426 CCAAAAGCCCGAGAAAAGAACCGAACGATACAAAACCAACAAAGCAAG 485  
||| |||: : : : : ||| ||| |||: : :  
Qy 133 SerLysSerGlnSerLysSerGlnThrArgAsn-AsnArgSerGlnSerAsnAlaAs 152  
|||: : : : : |||: : : : : |||: : : : :  
Db 486 AAAAAACACAAAAAAGAGAGACAAAAAACAACAGAAAAACACAAAGCGAGA 545  
||| |||: : : : : ||| ||| |||: : :  
Qy 152 nAlaHisPheThrSerLysArgLysAspThrLysArgArgGlnGluArgHisIleLyl 172  
||| |||: : : : : ||| ||| |||: : :  
Db 546 A-----CAACCGAGACACGCGCCACACAAAGAGAGACAGGAAAAACATGATGA 596  
||| |||: : : : : ||| ||| |||: : :  
Qy 172 sGlu---GluGlnAspLysGluMetThrSerAlaLysGlnHis 185  
||| |||: : : : : ||| ||| |||: : :  
Db 597 CGAGGAAAGAGGACAAAGAAACACAAAAACAGAAAAACAGAAACAGAAC 639  
||| |||: : : : : ||| ||| |||: : :

RESULT 34  
BE914647  
LOCUS BE914647 978 bp mRNA linear EST 29-SEP-2000  
DEFINITION 601665155F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3965365 5',  
mRNA sequence.

ACCESSION BE914647  
VERSION BE914647.1 GI:10413486  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 978)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: L1AM9136 row: m column: 14  
High quality sequence stop: 400.  
Location/Qualifiers  
1. .978  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3965365"  
/clone\_lib="NCI\_CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"

FEATURES  
source

source

1. .1122  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4644703"  
/clone\_lib="NIH\_MGC\_15"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'

```

/dev_stage="10 months, virgin"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 459 a 150 c 244 g 123 t 2 others
ORIGIN

```

Alignment Scores:  
Pred. No.: 0.822 Length: 978  
Score: 105.00 Matches: 19  
Percent Similarity: 58.46% Conservative: 19  
Best Local Similarity: 29.23% Mismatches: 27  
Query Match: 10.41% Indels: 0  
DB: 12 Gaps: 0

US-09-868-352-23 (1-192) x BE914647 (1-978)

Qy 114 AsnProArgAspIleLysPheAspGluArgProHisLysArgLysSerArgSer 133  
||| |||: : : : : |||: : : : : |||: : : : :  
Db 626 AAAAAAAGAGAGAGAGAAACAGCAAGAAACCAACCAAGAAAGAAAGA 685  
||| |||: : : : : ||| ||| |||: : :  
Qy 134 LysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAla 153  
||| |||: : : : : ||| ||| |||: : :  
Db 686 AAGCAAAAGAAAAAACAAGAAAGAAACAAACAAAAAAGAAACAAAAAGA 745  
||| |||: : : : : ||| ||| |||: : :  
Qy 154 HisPheThrSerLysArgLysAspThrLysArgGlnGluArgHisIleLysGlu 173  
|||: : : : : |||: : : : : |||: : : : :  
Db 746 GAAAGAGAGAAAGAAAAACGAGAGCAAGAAAGAAAGAAAGAAACACACAAAGA 805  
||| |||: : : : : ||| ||| |||: : :  
Qy 174 GluGlnAspLysGlu 178  
||| |||: : : : : ||| ||| |||: : :  
Db 806 AAGACGGAAGAAAGAA 820

RESULT 35  
BG468632  
LOCUS BG468632 1122 bp mRNA linear EST 21-MAR-2001  
DEFINITION 602510232F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:4644703 5',  
mRNA sequence.

ACCESSION BG468632  
VERSION BG468632.1 GI:13400902  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1122)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: NIH Intramural Sequencing Center  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CMI416 row: o column: 08  
High quality sequence stop: 404.  
Location/Qualifiers

1. .1122  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4644703"  
/clone\_lib="NIH\_MGC\_15"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'



```

Db 233 GTCAACGGCAGAAAAATCAGATTAGTTGAAGACAAGCCAGGTTCTAGGCGGCCCGCTCT 292
QY 126 -----HisLysArg-----
Db 293 TACTCCAGAAGCGGAGTCACTCAAGGTCCTCGCTCTGAAGCAGACATTCGCCGAAGAGC 352
QY 132 ArgSerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 151
Db 353 AGAAGCGGAGTGGTAGTAGTAAGCAAGCCATTCGAAGAGCGGCTCTCGATCCAGGTCA 412
QY 152 AsnAlaHisPheThrSerLysLysArg-----
Db 413 GGTTCCTCCACTCCGACAGAACGCCGCTGCGAGTCAGACGCCGACGCGCAGTAGAAG 472
QY 161 -----LysAspThrLysArgGlnGlnArgHisLysLysGlu 173
Db 473 GAGAAAGCGGAGCCGACCAAGCAACAGAGCGCGAGTCGC-----AGCCGC 523
QY 174 GluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Db 524 AGCCCTGACAAGAGCGCGCAGTAAGAGTAAGACCAC 559

RESULT 32
BJ111917
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .652
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_lib="yk1148h02"
/elegans_L1_stage"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pCC86"
BASE COUNT 192 a 135 c 171 g 154 t
ORIGIN
Alignment Scores:
Pred. No.: 0.532 Length: 652
Score: 105.00 Matches: 46
Percent Similarity: 35.53% Conservative: 24
Best Local Similarity: 23.35% Mismatches: 93
Query Match: 10.41% Indels: 34
DB: 13 Gaps: 5

```

```

US-09-868-352-23 (1-192) x BJ111917 (1-652)
QY 11 AsnTyrAsnLysTyrProGlnPheIleHisPheGluAsnIleValLysSerAsp 30
Db 32 AATTACAACAACAACAACATACCATCATCATCAACAATTTCCGAGAATCAAT 91
QY 31 AspIle-----
Db 92 GAGACCAAGTTCTCTCAGTCGAAGCACAGATTTTATTTCGAGGACTTCATGGTGATATT 151
QY 33 -----GluPheGlnLeuValIleAsnGlu-----LysSerAlaPheAspVal 46
Db 152 TGCACGGAAGAGATCAAGGAATATATTCGCGAAAGATTGGAAAAATCTCGTTTCGATTTC 211
QY 47 ThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAsp 66
Db 212 GTGAAAGTGGCTCAAGATAAGACAAGATATTCGTGGCTGTTTCGATTTCGAGAATCGTGAT 271
QY 67 TrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys 86
Db 272 GAGGCTAAGGAGTTTATGGAGACA-----TATTCGATCGTGATTCATGGGATGC 322
QY 87 AsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAla 106
Db 323 CGTTGTGATCTCTCTTGGTTCCTCGTGAT---ATACGGCTGTTATTCGCTATCAGCGTGCC 379
QY 107 TyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgProHis 126
Db 380 AAGCAGGTTTCGAAGCAACACAGTCAACGCCGCTCGTCTCATTCACAAGAGTCTTCG 439
QY 127 LysArg-----ArgLysSerArgSerLysSerGlnSerLysSer 140
Db 440 AAGAGATCTCGTCGCCACCAAGTCGAGAGAGGACGATCGAAGAGCGGATCTAGAAGCAGA 499
QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
Db 500 AGTCCGCTCAGAGAGCGGATCGAAGAGCGCTCAAGAAAGTGTGTGTAGATCAAGAGTCCCG 559
QY 161 LysAspThrLysArgGlnGluArgHisIleLysGluGlnAspLys 177
Db 560 GCATCTAGAAATGGAAGAGTCTCTCGGCATCGTAGTAGAAGTCAAGATCGA 610

RESULT 33
BJ1734757
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
BI734757 820 bp mRNA linear EST 20-SEP-2001
603356626F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5363670 5',
mRNA sequence.
BI734757 GI:15711770
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 820)
NIH-MGC http://mgi.nhl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11925 row: d column: 07
High quality sequence stop: 40.
Location/Qualifiers
1. .820
/organism="Mus musculus"
/db_xref="taxon:10090"

```

```

Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
Dy 157 TATCGAGATGCTCAACAGGGACGCAAAAC-----GAAGGAGTG 195
Qy 98 LysGluTyrCysAsnPhcGlyCysAlaTyrPheValLeuGlu-----Asn 112
Dy 196 ATTCAATTTGCTCTACTCTGATATGAAAAGAGCTTTGGAAAAGCTGGACGGAAGTAA 255
Qy 113 ProAsnProArgAspIleLysPheAspGluArgPro----- 125
Dy 256 GTCACGGCAGCAAAATCAGATTAGTGAAGACAGCCAGGTTCTAGCGCGCGCGTCT 315
Qy 126 -----HisLysArg----- 131
Dy 316 TACTCCAGAAGCGGAGTCACTCAAGGTCGCTCTCGAAGCAGACATTCCTCCGGAAGAGC 375
Qy 132 ArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 151
Dy 376 AGAAGCCGGAGTGGTAGCAGTAAAGACAGCCATTTCGAAGAGCGCGTCTCGATCCAGGTCA 435
Qy 152 AsnAlaHisPheThrSerLysLysArg----- 160
Dy 436 GTTCCCACTCCGAGCAGCAGCGGCGAGTGCAGTCAGACCGCGCAGCGAGTAAGAAC 495
Qy 161 -----LysAspThrLysArgArgGlnGluArgHisIleLysGlu 173
Dy 496 GAGAAAGCCGAGCGCCAGCAAGACAAAGAGAGCCGCGCTCGC-----AGCGCG 546
Qy 174 GluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Dy 547 AGCCCTGACAAGAGCGCGCAGTAAAGAGTAAGACCCAC 582

RESULT 31
BQ569343
LOCUS      BQ569343      615 bp      mRNA      linear      EST 19-JUN-2002
DEFINITION      g1126h02.y1 Mouse Organ of Corti cDNA pBluescript Mus musculus cDNA
                clone g1126h02 5', mRNA sequence.
ACCESSION      BQ569343
VERSION        BQ569343.1  GI:21472660
KEYWORDS        EST.
SOURCE          house mouse.
ORGANISM        Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 615)
Kachar,B.
EST analysis of gene expression in the mouse Organ of Corti at the
onset of hearing
Unpublished (2002)
Contact: Kachar,B.
Structural Cell Biology
National Institute of Deafness and other Communication Disorders
50/4249 South Drive, NIH, Bethesda, MD 20892-8027, USA
Tel: 301-402-1599
Fax: 301-402-1765
Email: kacharbenidcd.nih.gov
Plate: 126 row: h column: 02
Seq primer: M13Rp1 reverse primer (ABI).
                Location/Qualifiers
                1..615
                /organism="Mus musculus"
                /strain="BALB/c"
                /db_xref="taxon:10090"
                /clone="g1126h02"
                /clone_lib="Mouse Organ of Corti cDNA pBluescript"
                /sex="male and female"
                /dev_stage="Post natal day 5 to 13"
                /note="Organ: Organ of Corti; Vector: pBluescript; The
organ of Corti (OC) was fine dissected from a total of 386
OC as follows: 102 samples from post-natal (P) day 5; 72
from P6; 60 from P7; 46 from P8; 18 from P9; 20 from P10;
14 from P12 and 24 from P13. After killing animals by
cervical dislocation followed by decapitation, the bulla

```

was removed and opened in Leibowitz medium. The bony capsule of the cochlea was chipped away, stria vascularis and spiral ligament were removed and the sensory epithelium was carefully dissected out of the modiolus. Total RNA was extracted using the micro Fasttrack kit (catalog # K1593-02; Invitrogen, Carlsbad, CA), according to manufacturer's instructions. Reverse transcription and library construction were carried out with the Uni-Zap XR vector kit (catalog # 237211, Stratagene) and Uni-Zap XR GigaPack III Gold Cloning kit (catalog # 237612), both from Stratagene (La Jolla, CA, USA), according to manufacturer's instructions. Briefly: 1.5 ug mRNA was reverse transcribed using a hybrid oligo(dT) linker-primer that contains an Xho I site. First strand synthesis was primed with the linker-primer and transcribed using Moloney murine leukemia virus reverse transcriptase (MMuLV-RT) and 5-methyl dCTP. The second strand was synthesized with DNA polymerase and RNase H. Complementary DNA was blunt ended with Pfu DNA polymerase, ligated with EcoR I adapters in the presence of ligase and digested with Xho I. The cDNA was sequentially size fractionated over Pharmacia Size Sep400 (Pharmacia, Uppsala, Sweden) and Clontech Chroma Spin-1000 (Clontech, Palo Alto, CA) columns to enrich for cDNAs greater than 400bp and 1000 bp, respectively. The cDNA was then directionally ligated to the Uni-Zap XR vector, which had been predigested with EcoR I and Xho I. The phagemid was packaged with GigaPack III Gold and, upon titration on XlI Blue MRF' cells, the yield of the phage library was estimated to be 11,100,000 recombinants. Stratagene's ExAssist Interference resistance helper phage (catalogue # 211203) was adopted to rescue plasmid DNA from the phages. Upon plating of the rescued library, individual cDNA clones were selected and grown in 96-well, 2 ml growth plate. Plasmid DNA was purified from 200 ul of saturated culture with the Concert96(TM) plasmid purification kit (Invitrogen, Carlsbad, CA) as instructed by the manufacturer. ESTs from the 5' end of the cDNA clones were generated with the universal M13 reverse primer (CAGGAACAGCTATGACC) and 25% strength BigDye terminator sequencing chemistry (Applied Biosystems, Foster City, CA). Sequencing reactions were performed on MJ Tetrad thermal cyclers (MJ Research, Waltham, MA), and analyzed on 3700 automated capillary sequencers using POP5 polymer (Applied Biosystems, Foster City, CA). The frequency distribution of the library is as follows: 72% of genes have 1 copy; 14.3% 2; 12% 3-10; 1.4% 11-50 and 0.1% 51-150. As to gene function, 45% of genes are present in GenBank and have known function; 23% have hits in GenBank, but do not have assigned function; 12% are uncharacterized ESTs and 20% are unidentified."

BASE COUNT 189 a 161 c 176 g 89 t

ORIGIN

Alignment Scores:  
 Pred. No.: 0.5 Length: 615  
 Score: 105.00 Matches: 40  
 Percent Similarity: 36.84% Conservative: 16  
 Best Local Similarity: 26.32% Mismatches: 42  
 Query Match: 10.41% Indels: 54  
 DB: 14 Gaps: 6

US-09-868-352-23 (1-192) x BQ569343 (1-615)

Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97  
 Dy 134 TATCGAGATGCTCAACAGGGACGCAAAAC-----GAAGGAGTG 172  
 Qy 98 LysGluTyrCysAsnPhcGlyCysAlaTyrPheValLeuGlu-----Asn 112  
 Dy 173 ATTCAATTTGCTCTACTCTGATATGAAAAGAGCTTTGGAAAAGCTGGACGGAAGTAA 232  
 Qy 113 ProAsnProArgAspIleLysPheAspGluArgPro----- 125

```

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       NIH-MGC http://mgs.nci.nih.gov/
JOURNAL     National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT     Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLC1817 row: h column: 20
            High quality sequence stop: 101.

FEATURES    Location/Qualifiers
            1..731
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4909507"
                /clone_lib="NIH_MGC_42"
                /tissue_type="epithelioid carcinoma cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
                Site:2: EcoRI; CDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
                for average insert size 1.8kb. Library constructed by Ling
                Hong in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC library. |"
            439 a 131 c 125 g 36 t
            BASE COUNT 439 a 131 c 125 g 36 t
            ORIGIN

Alignment Scores:
Pred. No.: 0.535 Length: 731
Score: 105.50 Matches: 34
Percent Similarity: 50.94% Conservative: 20
Best Local Similarity: 32.08% Mismatches: 38
Query Match: 10.46% Indels: 15
DB: 12 Gaps: 4

US-09-868-352-23 (1-192) x BG830845 (1-731)

QY 79 LysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLys 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 AAAAAACGAAGACCGAGACACATAGATCAAGAACACACAAAAAGAAACAAAAACAAA 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 99 GluTyrCysAsnPhgGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 322 GAGAAAGGAACAAAGGC-----AAAACGGAGCAGCAGACCAACAGAAACAGAAATA 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 119 LysPheAspGluArgProHisLys-----ArgArgLysSerArgSerLysSerGln 136
    ::||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 -----CAAAAGGAAGCACAAGGAAGACAAAATAAAAAACACAGACAAAACGCACAC 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 SerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThr 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 AAGAAAAAACAGACGCGACACAAAGAAACAAAGAAACACACAGGAAGCAAAAAACAA 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 157 SerLysLysArg-----LysAspThrLysArgArgGlnGluArgHisIleLys 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 GAGAAAGAAACGGAACCAACCAAGAAACCAACCAACCAACCAACCAACCAACCAACAA 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 173 GluGluGlnAspLysGlu 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 GAACACAAGACAAGAA 554

```

```

RESULT 30
LOCUS    BG086260
DEFINITION H3123G04-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
          H3123G04 5', mRNA sequence.
ACCESSION BG086260
VERSION    BG086260.1 GI:12568824
KEYWORDS   EST.
SOURCE     house mouse.
          Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
          1 (bases 1 to 613)
          Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
          ,T.S., Carter,M.G. and Ko,M.S.H.
          Verification and initial annotation of NIA mouse 15K cDNA clone set
          Unpublished (2001)
          Other_ESTs: H3123G04-3
          Contact: George J. Kargul
          Laboratory of Genetics
          National Institute on Aging/National Institutes of Health
          333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
          Email: cdna@igsun.grc.nia.nih.gov
          This clone set has been freely distributed to the community. Please
          visit http://igsun.grc.nia.nih.gov/cDNA/15k.html for details.
          Plate: H3123 row: G column: 04
          Seq primer: -21M13 Reverse
          High quality sequence stop: 613
          POLYA=No.

FEATURES    Location/Qualifiers
            1..613
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="niaEST:H3123G04-5"
                /db_xref="taxon:10090"
                /clone="H3123G04"
                /clone_lib="NIA Mouse 15K cDNA Clone Set"
                /sex="Clones arrayed from a variety of cDNA libraries"
                /dev_stage="Clones arrayed from a variety of cDNA
                libraries"
                /lab_host="DH10B"
                /note="Vector: pSPORT1; Site:1: SalI; Site:2: NotI; This
                clone is among a rearrayed set of 15,247 clones from 11
                embryo cDNA libraries (including preimplantation stage
                embryos from unfertilized egg to blastocyst, embryonic
                part of E7.5 embryos, extraembryonic part of E7.5 embryos
                , and E12.5 female mesonephros/gonad) and one newborn
                ovary cDNA library. Average insert size 1.5 kb. All
                source libraries are cloned unidirectionally with Oligo(dT
                )-Not primers. References include: (1) Genome-wide
                expression profiling of mid-gestation placenta and embryo
                using a 15,000 mouse developmental cDNA microarray, 2000,
                Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                Large-scale cDNA analysis reveals phased gene expression
                patterns during preimplantation mouse development, 2000,
                Development, 127: 1737-1749; (3) Genome-wide mapping of
                unselected transcripts from extraembryonic tissue of
                7.5-day mouse embryos reveals enrichment in the t-complex
                and under-representation on the X chromosome, 1998, Hum
                Mol Genet 7: 1967-1978."
            191 a 149 c 177 g 96 t
            BASE COUNT 191 a 149 c 177 g 96 t
            ORIGIN

Alignment Scores:
Pred. No.: 0.498 Length: 613
Score: 105.00 Matches: 40
Percent Similarity: 36.84% Conservative: 16
Best Local Similarity: 26.32% Mismatches: 42
Query Match: 10.41% Indels: 54
DB: 12 Gaps: 6

US-09-868-352-23 (1-192) x BG086260 (1-613)

```

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LAM9767 row: c column: 11  
 High quality sequence stop: 566.  
 Location/Qualifiers  
 1. 1728  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4206274"  
 /lab\_host="NCI-CGAP\_Co24"  
 /note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI-CGAP Library."  
 BASE COUNT 599 a 398 c 518 g 213 t  
 ORIGIN

## FEATURES

## source

1. 1728  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4206274"  
 /lab\_host="NCI-CGAP\_Co24"  
 /note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI-CGAP Library."  
 BASE COUNT 599 a 398 c 518 g 213 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.19 Length: 1728  
 Score: 106.00 Matches: 27  
 Percent Similarity: 44.57% Conservative: 14  
 Best Local Similarity: 29.35% Mismatches: 31  
 Query Match: 10.51% Indels: 20  
 DB: 12 Gaps: 2

US-09-868-352-23 (1-192) x BF578008 (1-1728)

Qy 112 AsnProAsnProArgAspPheAspGluArgProHisLysArgArgLysSer 131  
 Db 1302 AACCCAGAGCAAGACACATAGCAACAGCAGCAGAGAGCGGAGAACGA 1361  
 Qy 132 ArgSerLysSerGlnSerLysSerGln----- 141  
 Db 1362 AGAACGAGAGAGCGCAGCAACAGACACAAAGAGAGAAAGCGGAACGCACAG 1421  
 Qy 142 -----ThrArgAsnAsnArgSerGlnSerAsnAlaAsn 152  
 Db 1422 ACAGGAAGAGAGAAACAGACAAAGACAGCGGAGAGAACACAGCAAAACAGAAC 1481  
 Qy 153 AlaHisPheThrSerLysArgLysAspThrLysArgArgGlnGlu---ArgHisIle 171  
 Db 1482 ACAGCAGCAACAGCAACGAGAAACAGACAGACAGCGAGCAAGCGGTAGGAAGGT 1541  
 Qy 172 LysGluGluGlnAspLysGluMetThrSerAlaLys 183  
 Db 1542 GAGAGGAAGCAAGCAAGAGCGCCAGGAACAAAG 1577

## RESULT 28

BG288695 1858 bp mRNA linear EST 21-FEB-2001  
 LOCUS 602385507F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4514496 5',  
 DEFINITION mRNA sequence.

ACCESSION BG288695

VERSION BG288695.1 GI:13043791

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 1858)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LAM10402 row: n column: 01

High quality sequence start: 4

High quality sequence stop: 542.

Location/Qualifiers

1. 1858  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4514496"  
 /clone\_lib="NIH\_MGC\_93"  
 /tissue\_type="transitional cell papilloma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
 BASE COUNT 703 a 375 c 537 g 243 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.29 Length: 1858  
 Score: 106.00 Matches: 30  
 Percent Similarity: 42.42% Conservative: 26  
 Best Local Similarity: 22.73% Mismatches: 40  
 Query Match: 10.51% Indels: 36  
 DB: 12 Gaps: 4

US-09-868-352-23 (1-192) x BG288695 (1-1858)

Qy 77 PheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyr 96  
 Db 910 TTCATCTAAGCTGTAAGCAGCAACGAAACGAGCAGCAACAGCAGCAGCAGCAGCAG 969  
 Qy 97 IleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArg 116  
 Db 970 CAGGAACACAGGTGC-----AGCAAGACACACACAGA 1002  
 Qy 117 AspileLysPheAspAspGluArgProHis-----LysArgArgLysSerArgSer 133  
 Db 1003 CGAGCAGCAGAGAGCGCAAGAACACCACTAACACAGCAGCAGCAAGGAAGCAAGAG 1062  
 Qy 134 LysSerGlnSerSerLysSerGln-----ThrArgAsnAsnArgSer 147  
 Db 1063 ATGACCCAGACAGAGAAAGAAAGACAGCAGCCCTAACAGCAGACAGAAAGAGCCACC 1122  
 Qy 148 GlnSerAsnAlaAsnAlaHisPheThrSerLysLys----- 159  
 Db 1123 AAGAGCAAGCAGGAATACACAGCAGCAGAGTCGAGAGAGCAGCAGCATGCAAGCAGCAG 1182  
 Qy 160 -----ArgLysAspThrLysArgArgGlnGluArg 169  
 Db 1183 GACCAACACAGAGAGAGCGAGCCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1242  
 Qy 170 HistLysGluGluGlnAspLysGluMetThrSer 181  
 Db 1243 CCGAGCAGCAGAGGGGAGCAGCCGCAAGACACAGCA 1278

## RESULT 29

BG830845

LOCUS 602767576F1 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:4909507 5',

DEFINITION mRNA sequence.

ACCESSION BG830845

VERSION BG830845.1 GI:14178432

KEYWORDS EST.



Best Local Similarity: 29.81% Mismatches: 33  
 Query Match: 10.51% Indels: 19  
 DB: 13 Gaps: 2

US-09-868-352-23 (1-192) x BI552711 (1-878)

Qy 82 SerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCys 101  
 ||| :|||:|||||:||||| :|||:|||||:|||||  
 Db 622 AGCAACAAAGAAAGAAACACACAGAAACAAAGAAAGACAGACAAAGAA----- 675  
 Qy 102 AsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPheAsp 121  
 :|||:|||||:|||||:|||||:|||||:|||||  
 Db 676 -----AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 693  
 Qy 122 AspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGln 141  
 ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 694 AAGAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 753  
 Qy 142 ThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrLysLysArgLys 161  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 754 AAAAAACAAATAATAAAAAAAACCCACACACAGAAAAAAGAAAGAAAGAAAGAAAGAA 813  
 Qy 161 sAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSe 181  
 ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 814 A-----AAAAGACAGACAGAGAAACGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 867  
 Qy 181 rAlaLysGln 184  
 |||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 868 AAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 877

RESULT 24  
 BG490668 1148 bp mRNA linear EST 27-MAR-2001  
 LOCUS 602520145F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4638802 5',  
 DEFINITION mRNA sequence.

ACCESSION BG490668  
 VERSION BG490668.1 GI:13452178  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1148)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DTF/Gazdar  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Cloned through the I.M.A.G.E. Consortium/LLNL at:  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LICM1401 row: i column: 11  
 High quality sequence stop: 150.

FEATURES  
 source

1..1148  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4638802"  
 /clone\_lib="NIH\_MGC\_18"  
 /tissue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAAGAG(G). Library constructed by Ling Hong in the  
 Laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC Library."

BASE COUNT 699 a 134 c 289 g 26 t  
 ORIGIN

Alignment Scores:  
 Pred. NO.: 0.771 Length: 1148  
 Score: 106.00 Matches: 28  
 Percent Similarity: 41.75% Conservative: 15  
 Best Local Similarity: 27.18% Mismatches: 38  
 Query Match: 10.51% Indels: 22  
 DB: 12 Gaps: 2

US-09-868-352-23 (1-192) x BG490668 (1-1148)

Qy 111 GluAsnProAsnProArgAspIleLysPheAspGluArgProHisLysArgArgLys 130  
 ||| :|||:|||||:|||||:|||||:|||||:|||||  
 Db 824 GAAACACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 883  
 Qy 131 SerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsn 150  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 884 GAAAGGAATAAGGGCAACGAGAAAAAGCAACAGAGAAG---AAACGAGAGAAAAAGAAA 940  
 Qy 151 AlaAsnAlaHisPheThrSerLysLysArgLys----- 161  
 ||| :|||:|||||:|||||:|||||:|||||:|||||  
 Db 941 CAGAACAAACAGAAACAGAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1000  
 Qy 162 -----AspThrLysArgArgGlnGluArg 169  
 |||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 1001 GAGAGAAACACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1060  
 Qy 170 HisIleLysGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuLeuPheVal 189  
 ||| :|||:|||||:|||||:|||||:|||||:|||||  
 Db 1061 CACAAAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1120  
 Qy 190 ArgLysAsn 192  
 |||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 1121 CGAAGAGAAC 1129

RESULT 25  
 BF796631 1260 bp mRNA linear EST 12-JAN-2001  
 LOCUS 602259829F1 NIH\_MGC\_85 Homo sapiens cDNA clone IMAGE:4343119 5',  
 DEFINITION mRNA sequence.

ACCESSION BF796631  
 VERSION BF796631.1 GI:12101685  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1260)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Louis Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Cloned through the I.M.A.G.E. Consortium/LLNL at:  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM9959 row: i column: 08  
 High quality sequence stop: 313.

FEATURES  
 source

1..1260  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4343119"  
 /clone\_lib="NIH\_MGC\_85"  
 /tissue\_type="lymphoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lymph; Vector: pCMV-SPORT6; Site\_1: NotI;







```

QY 174 GluGlnAspLysGlu-----MetThrSerAlaLysGln 184
      :::::::::::::::
Db 1033 AAAGAGAACAAAGCAGCAGCACACAAACAGGAGGAAACAA 1074

RESULT 19
AA838525      511 bp      mRNA      linear      EST 18-MAR-1998
LOCUS      oe90h12.s1 NCI_CGAP_Col2 Homo sapiens cDNA clone IMAGE:1418951,
DEFINITION      mRNA sequence.
ACCESSION      AA838525
VERSION      AA838525.1 GI:2913324
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 511)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL      Tumor Gene Index
COMMENT      Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Sequencing by: Greg Lennon, Ph.D.
Clone Distribution: NCI-CGAP clone distribution
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 666 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 54.

FEATURES
    source
        1..511
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1418951"
            /clone_lib="NCI_CGAP_Col2"
            /sex="mixed"
            /tissue_type="colon tumor"
            /lab_host="SOLR (kanamycin resistant)"
            /note="Organ: colon; Vector: Bluescript SK-; Site_1: EcoRI
            ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
            dr. Pooled colon tumors. 5' adaptor sequence: 5'
            GAATTCGGCAGCAG 3' 3' adaptor sequence: 5'
            CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

BASE COUNT      267 a 104 c 79 g 61 t
ORIGIN

Alignment Scores:
Pred. No.:      0.288      Length:      511
Score:          106.50      Matches:      27
Percent Similarity: 49.00%      Conservative: 22
Best Local Similarity: 27.00%      Mismatches: 36
Query Match:      10.56%      Indels:      15
DB:              9      Gaps:      3

US-09-868-352-23 (1-192) x AA838525 (1-511)

QY 104 GlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspGlu 123
      :::::::::::::::
Db 122 GGCTGTAGT-----AAGCCTCAAGGCCACTCTCAACATGAGA-----GAAGAA 169

QY 124 ArgProHisLysArgArgLysSerArgSerLysSerGlnSerLysSer----- 140
      :::::::::::::::
Db 170 AAACCATACAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 229

QY 141 -----GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsn 152
      :::::::::::::::
Db 230 ATACACAGACTGCTACAAATACACAAACAAAGAAACAGACCAACCAACAAACAAAC 289
      :::::::::::::::

```

```

QY 153 AlaHisPheThrSerLysLysArgLysAspThrLysArgGlnGluArgHisIleLys 172
      :::::::::::::::
Db 290 ATACACCACACAAATAAGAGAGACACACACACCAACCGAAACAAACCAACCAAAA 349

QY 173 GluGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuPheValArgLysAsn 192
      :::::::::::::::
Db 350 AAACAAACCAAGAGAGAAACAGAAACAAACAAACAAACAAACAAACACACCAACAT 409

RESULT 20
BB619189      966 bp      mRNA      linear      EST 26-OCT-2001
LOCUS      BB619189
DEFINITION      BB619189 RIKEN full-length enriched, 8 days embryo Mus musculus
ACCESSION      BB619189
VERSION      BB619189
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 966)
AUTHORS      Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
            Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
            RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
            Unpublished (2001)
TITLE      Contact: Yoshihide Hayashizaki
JOURNAL      Laboratory for Genome Exploration Research Group, RIKEN Genomic
COMMENT      Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yananaka,I., Aizawa
            K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
            Location/Qualifiers
                1..966
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="5730499p16"
                    /clone_lib="RIKEN full-length enriched, 8 days embryo"
                    /sex="mixed"
                    /dev_stage="8 days embryo"

```

[illegible]

## Alignment Scores:

Pred. No.: 0.418 Length: 1256  
 Score: 109.00 Matches: 37  
 Percent Similarity: 41.55% Conservative: 22  
 Best Local Similarity: 26.06% Mismatches: 52  
 Query Match: 10.80% Indels: 31  
 DB: 12 Gaps: 4

US-09-868-352-23 (1-192) x BG121867 (1-1256)

QY 68 GlyAsnGluLeuArgLeuArgGlyPheTyrLysAspAla----- 81  
 Db 724 GGAGAGGAGCCCGCAGACGACGCGAGCGAGCGCTACAGGAAAGGCCCTTGG 783  
 QY 82 ---SerThrIleArgLysAsnSerArgIleSer----- 91  
 Db 784 GAATCAACCATCTTAGAAGGGAGCAGACACAAACCGGAGGCAACCATCCGGAGTATCGAG 843  
 QY 92 -----ArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAla 106  
 Db 844 CCCACATAGTCCCTAACATACAGAACACACACAGCTCGTGGCTTCATGACGATA 903  
 QY 107 TyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspGluArgProHis 126  
 Db 904 TTGTATCTCTTCACTGCTCTAACGACGACGACCGCAACATGAGACACACAGTACAG 963  
 QY 127 LysArgArgLysSerArgSerLysSerGln-----SerSerLysSerGlnThr 142  
 Db 964 AGAAGAAGAGGATCAACAGAAAGAGCAGACGAGGATAGCGAGCAAGACGAGCAGC 1023  
 QY 143 ArgAsn-----AsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSer 157  
 Db 1024 GAGACGAAACCCAAAGAAACAAAGGCAAGCAACAGCAAGCAAGCAAGCAAGCA 1083  
 QY 158 LysLysArgLysAspThrLys-ArgArgGlnGluArgHisIleLysGluGluGlnAsp 177  
 Db 1084 AGAAGAAGAAGAACTAAACAGAAACAGACAGACACATAAAGAGAAACAACAAGAAG 1143  
 QY 177 sglu 178  
 Db 1144 AGAA 1147

RESULT 16  
 BM880093  
 LOCUS ku04h04.y1 Strongyloides ratti PA female naive pAMP1 v1  
 DEFINITION Strongyloides ratti cdna 5', mRNA sequence.  
 ACCESSION BM880093  
 VERSION BM880093.1 GI:19252760  
 KEYWORDS EST.  
 SOURCE Strongyloides ratti.  
 ORGANISM Strongyloides ratti  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Panagrolaimoidea; Strongyloidea; Strongyloides.

REFERENCE 1 (bases 1 to 471)  
 AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
 TITLE The Washington Univ. Nematode EST Project, 1999  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: McCarter JP  
 The Washington Univ. Nematode EST Project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 The library was constructed by Brandi Chiapelli and Dr. James

McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Dissected nematode tissues were provided by Dr. Alan Scott. (ascott@hsp.edu) of the School of Public Hygiene and Public Health at John Hopkins University in Baltimore, MD.  
 High quality sequence stop: 395.

FEATURES  
 source

1. 471  
 Location/Qualifiers  
 /organism="Strongyloides ratti"  
 /db\_xref="taxon:34506"  
 /clone\_lib="Strongyloides ratti PA female naive pAMP1 v1"  
 /dev\_stage="parasitic adult females"  
 /lab\_host="DH10B"  
 /note="Vector: pAMP1 (Gibco); Site.1: NotI; Site.2: SalI;  
 The library was constructed by Claire Murphy, Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Parasitic adult females were collected from naive animals and provided by Dr. Mark Viney of Bristol, UK."

BASE COUNT 190 a 58 c 108 g 115 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 0.208 Length: 471  
 Score: 107.50 Matches: 36  
 Percent Similarity: 45.64% Conservative: 32  
 Best Local Similarity: 24.16% Mismatches: 62  
 Query Match: 10.65% Indels: 20  
 DB: 14 Gaps: 3

US-09-868-352-23 (1-192) x BM880093 (1-471)

QY 46 ValThrValPheGlyClnArgPheSerGluIleLeuLysTyrAspPheIleValGly 65  
 Db 52 ATAAATATGATGGGATATCAATATTCAAGTCAATGCTTCTCAAAATGCCTATACAGGT 111  
 QY 66 -----AspTyrGlyAsnGluGlnLeuArgLeuArg 75  
 Db 112 GGTGTTATTATCAATCTTACTATGCAAAATGTTGTCAAAATGGCTGATCAATTTAGT 171  
 QY 76 GlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAsp 95  
 Db 172 GGTGTTACGAGAGTGGAAGTAGAAGTGGTGAAGATCAAGATCTCTAGTTCTTCTTCA 231  
 QY 96 TyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnPro 115  
 Db 232 TATCTTCTTAGTGGGAGT-----CGAAGAAGTAGATCATCT 270  
 QY 116 ArgAspIleLysPheAspAspGluArgProHisLysArgArgLysSerArgSerLysSer 135  
 Db 271 AGGATAGTAGTAGTATCACTAGAGGATCGAAGAAAGATCATCTAGAAAAAGT 330  
 QY 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155  
 Db 331 AGG-----AAAAATAGGAAAAAGATTTCTAGAGGAGATCTAGATCTAGAAAGTTACTCT 384  
 QY 156 ThrSerLysArgLysArgThrLysArgGlnGluArgHisIleLysGluGluGln 175  
 Db 385 TCTGATAGAAAAAGAAAGATGATCATATAAAAAAG-GATAGGAGTGGAAAGGAGAA 443  
 QY 176 AspLysGluMetThrSerAlaLysGln 184  
 Db 444 AAAAAGAAAGATAAAGTAGTAAAAAG 470

RESULT 17  
 BF672136  
 LOCUS 602152592F1 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:4293575 5',  
 DEFINITION

1527 bp mRNA linear EST 21-DEC-2000  
 BF672136  
 LOCUS 602152592F1 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:4293575 5',  
 DEFINITION

```

Db 133 ATAGAGTGTGATGTGTGTTGGTGTCTGTTACCGTGTGATGAATGAAGAGAGAGAGAG 192
      ::::: ||| ||||| ||| ::::: |||
Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
      ||| ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 TAT-----AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 222
      ::::: ||| ||||| ||| ::::: |||
Qy 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAsp 117
      ::::: ||| ||||| ||| ::::: |||
Db 223 TATAAATAT-----ATATAACATGATGAAGTAAATAATGAGAAGT 258
      ::::: ||| ||||| ||| ::::: |||
Qy 118 IleLysPheAspAsp-----GluArgProHisLysArgArgLysSerArgSerLysSer 135
      ||| ::::: ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 259 GAAAAAGAGAATAATATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 318
      ::::: ||| ||||| ||| ::::: |||
Qy 136 GlnSerSerLys-----SerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAla 153
      ::::: ||| ||||| ||| ::::: |||
Db 319 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 378
      ::::: ||| ||||| ||| ::::: |||
Qy 154 HisPheThrSerLysLysArgLysAspThrLysArgArgGlnGlu---ArgHisIleLys 172
      ||| ||||| ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 CACACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 438
      ::::: ||| ||||| ||| ::::: |||
Qy 173 GluGluGlnAspLysGluMetThrSerAlaLys 183
      ||| ::::: ||||| ||| ::::: |||
Db 439 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471
      ::::: ||| ||||| ||| ::::: |||

RESULT 14
BF684615 814 bp mRNA linear EST 22-DEC-2000
LOCUS 602140956F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4302205 5',
DEFINITION mRNA sequence.
ACCESSION BF684615.1 GI:11970023
VERSION BF684615.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 814)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM163 row: p column: 14
High quality sequence stop: 382.
Location/Qualifiers
1. .814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4302205"
/clone_lib="NIH_MGC_46"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
357 a 202 c 154 g 101 t

FEATURES
source
1. .814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4302205"
/clone_lib="NIH_MGC_46"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 357 a 202 c 154 g 101 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 0.262 Length: 814
Score: 109.00 Matches: 25
Percent Similarity: 52.22% Conservative: 22
Best Local Similarity: 27.78% Mismatches: 33
Query Match: 10.80% Indels: 10
DB: 12 Gaps: 2

US-09-868-352-23 (1-192) x BF684615 (1-814)
Qy 112 AsnProAsnProArgAspIleLysPheAspAspGluArgProHisLysArgArgLysSer 131
      ::::: ||| ||||| ||| ::::: |||
Db 370 AGCCCAACTGAGTGTACTCTACCTCTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 429
      ::::: ||| ||||| ||| ::::: |||
Qy 132 -----ArgSerLysSerGlnSerSerLysSerGlnThr 142
      ::::: ||| ||||| ||| ::::: |||
Db 430 AAGACACACGAGACACACACACACAGTAAGTAAGAAACACGAAAGAAAGAAAGAA 489
      ::::: ||| ||||| ||| ::::: |||
Qy 143 ArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHis---PheThrSerLysLysArgLys 161
      ::::: ||| ||||| ||| ::::: |||
Db 490 AAAAAACAAAAGAAAAACACGACACAGACACACAGCAGCAGCAGCAGCAGCAGCAG 549
      ::::: ||| ||||| ||| ::::: |||
Qy 162 AspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSer 181
      ::::: ||| ||||| ||| ::::: |||
Db 550 GAAACAGAAAGAAAGTAAGAAACACATATAAGTAAAGTAAAGTAAAGTAAAGTAAAG 609
      ::::: ||| ||||| ||| ::::: |||
Qy 182 AlaLysGlnHisLeuLeuPheValArgLys 191
      ::::: ||| ||||| ||| ::::: |||
Db 610 AGACAGCAACATATAAGAAAGAAACTACAGAAAA 639
      ::::: ||| ||||| ||| ::::: |||

RESULT 15
BG121867 1256 bp mRNA linear EST 30-JAN-2001
LOCUS 602350882F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4445341 5',
DEFINITION mRNA sequence.
ACCESSION BG121867
VERSION BG121867.1 GI:12615376
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1256)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10223 row: d column: 22
High quality sequence stop: 571.
Location/Qualifiers
1. .1256
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4445341"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
439 a 303 c 315 g 199 t

FEATURES
source
1. .1256
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4445341"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 439 a 303 c 315 g 199 t
ORIGIN

```



```

Qy 153 AlaHisPheThrSerLysLysArgLysAspThrLysArgGlnGlnLysHisLysLeu 172
    ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 593 AAA-----ACAGAGAAAAAACTAAAGACACAAAAACACAGACGAAAGAG 646
    ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 173 GluGlnAspLysGluMet-----ThrSerAla 182
    :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 647 CAAGAAGACGTAAGAGATCAGAACGCAAAAAAGAGACCAAGCAAGCGACAAACACA 706
    :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Qy 183 LysGlnHisLeuLeuPheValArgLysAsn 192
    ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 AAAAACCAACACATAAAAGAGACGAAC 736
    ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
LOCUS AL640539 505 bp mRNA linear EST 12-DEC-2001
DEFINITION AL640539 XGC-neurula Silurana tropicalis cDNA clone TNeu04k13 5',
    mRNA sequence.
ACCESSION AL640539
VERSION AL640539.1 GI:16792670
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
    xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 505)
AUTHORS Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2001 (10_2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Huckle E
    Sanger Centre
    Hinxton, Cambridgeshire, CB10 1SA, UK
    Email: tropesanger.ac.uk
    Sanger Xenopus tropicalis EST project 2001
    TROPICALIS_SEQUENCE_ID: TNeu04k13.sp6
    Sequencing primer: SP6
    This sequence is from a Xenopus Gene Collection (XGC) library
    constructed by Aaron M. Zorn.
FEATURES
    source
        1..505
            location/Qualifiers
                /organism="Silurana tropicalis"
                /db_xref="taxon:8364"
                /clone="TNeu04k13"
                /clone_lib="XGC-neurula"
                /dev_stage="neurula"
                /lab_host="Escherichia coli DH10B"
                /note="vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                    was oligo dt primed from Sug of poly A+ RNA from neurula.
                    EcoRI-NotI cut cDNA was then ligated into pCS107 with
                    EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT 105 a 178 c 145 g 76 t
ORIGIN
Alignment Scores:
    0.14 Length: 505
    109.50 Matches: 36
    Percent Similarity: 46.03%
    Best Local Similarity: 28.57%
    Query Match: 12
    Indels: 9
    Gaps: 3
    DB:

US-09-868-352-23 (1-192) x AL640539 (1-505)

Qy 68 GlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsn 87
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34 GGAGAGCGGGGCTTCGCTTCGTCGTCACAGCAAGCGCGATGCCGAAGACGCGAT 93
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 88 SerArg-IleSer-----ArgLeuGluAspTyrIleLysGluTy 100
    ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GGACCCATGATGGGGCGGCTGCTGGACGGCGGAGCTGAGGTGCAGATGGCCCGGTA 153
    ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 100 rCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArg-----As 117

```

```

|||||: :|||: |||||
Db 154 CTGTTCTTCTCCGATTCCCAACACGGAAGAGAGGGCCCTCTCTCCGACATACGGGA 213
    |||||: :|||: |||||
Qy 117 pileLysPheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSe 137
    |||||: :|||: |||||
Db 214 CTAGCGCGCGGAGACAGAAAGCCAGGAGCGCGTCCGACGCGTTCAGAACGACAG 273
    |||||: :|||: |||||
Qy 137 r---SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheTh 156
    |||||: :|||: |||||
Db 274 CAGTCAAGGTCCCGCAGCAGGTCTCGCTACAGCGCTCAAGTCCCGCTCTCCGACAG 333
    |||||: :|||: |||||
Qy 156 rSerLysLysArgLysAspThrLysArgArgGlnGlnLysHisLysGluGlnGlnAs 176
    |||||: :|||: |||||
Db 334 GTCCCGCAGCCCGTCCAGTCAAAAGTCTCGCTCGCGCAAGGAGATCCCAAGTCTAAGTCATC 393
    |||||: :|||: |||||
Qy 176 pLysGluMetThrSer 181
    |||
Db 394 GTCCGCGTCCAGGTCT 409
    |||

RESULT 11
LOCUS BG259084 1005 bp mRNA linear EST 13-FEB-2001
DEFINITION 602379137F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4509800 5',
    mRNA sequence.
ACCESSION BG259084
VERSION BG259084.1 GI:12768887
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1005)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
    Email: cgabbs-r@mail.nih.gov
    Tissue Procurement: ATCC
    CDNA Library Preparation: Life Technologies, Inc.
    CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
    DNA Sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    http://image.llnl.gov
    Plate: L1AM10390 row: j column: 09
    High quality sequence stop: 438.
    Location/Qualifiers
        1..1005
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4509800"
            /clone_lib="NIH_MGC_92"
            /tissue_type="embryonal carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 2.5 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT 467 a 202 c 211 g 125 t
ORIGIN

Alignment Scores:
    0.292 Length: 1005
    109.50 Matches: 25
    Percent Similarity: 56.16%
    Best Local Similarity: 34.25%
    Query Match: 10.85%
    Indels: 5
    Gaps: 2
    DB:

US-09-868-352-23 (1-192) x BG259084 (1-1005)

Qy 113 ProAsnProArgAspIleLysPheAspGluArgProHisLysArgArgLysSerArg 132

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaest:L0230G12-3"
/db_xref="taxon:10090"
/clone="L0230G12"
/clone_lib="NIA Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/notes="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
7 Newborn Ovary. The double-stranded cDNA was synthesized
by Gibco's kit with an oligo(dT) primer [NotI
primer-adaptor from Gibco/BRL]
[5'-GGACTAGTCTAGATCGGAGCGGCCCTTTT-3'] from
2.56ug of total RNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
lone-linker LL-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Yulan Piao."
BASE COUNT      114 a 154 c 102 g 181 t
ORIGIN

Alignment Scores:
Pred. No.:      0.121      Length:      551
Score:          110.50     Matches:      30
Percent Similarity: 49.44%  Conservative: 14
Best Local Similarity: 33.71% Mismatches:    34
Query Match:      10.95%   Indels:       11
DB:               10      Gaps:         3

US-09-868-352-23 (1-192) x AW553719 (1-551)

QY 103 PheGlyCysAla-----TyrPheValLeuGluAsnProAsnPro 115
Db 384 TTTAATTGTTCTGAGCTTACGCTGCAAGTGGTTTTTCTCCAGAGTCTGGAAGAGC 325

QY 116 ArgAspIleLysPheAspGluArgProHisLysArgArgLysSerArgSerLysSer 135
Db 324 AATGAAGGAAGCGGTGATCTGTTCCACACTCGCTCAAGTCCAGTCT----- 271

QY 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
Db 270 ---AGTCAAAATCCCACCTCTAGAAAGAGAGATCATCAGTCAAGACACAGATCCCATAT 214

QY 156 ThrSerLysArg---LysAspThrLysArgArgGlnGluArgHisIleLysGluGlu 174
Db 213 AGTCTACGCTCAAGACAGAAGATAGATAGATCCAAAGAGCCCAACACAAAGACGCTCT 154

QY 175 GlnAspLysGluMetThrSerAlaLys 183
Db 153 AAGTCTAGGAGAGCGGAAGTCAAGG 127

RESULT 9
BE899526
LOCUS      BE899526      999 bp      mRNA      linear      EST 29-SEP-2000
DEFINITION BE899526 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3951885 5',
            mRNA sequence.
ACCESSION BE899526
VERSION    BE899526.1 GI:10367116
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LLCM820 row: k column: 22
High quality sequence start: 5
High quality sequence stop: 361.

```

#### FEATURES

```

Location/Qualifiers
1..999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3951885"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      459 a 203 c 253 g 83 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      0.229      Length:      999
Score:          110.50     Matches:      47
Percent Similarity: 38.42%  Conservative: 26
Best Local Similarity: 24.74% Mismatches:    66
Query Match:      10.95%   Indels:       51
DB:               12      Gaps:         7

```

```

US-09-868-352-23 (1-192) x BE899526 (1-999)

```

```

QY 31 AspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50
Db 236 GATGCTGAGGAAGATTGTGTGGTCCAGCCAGCAGTGGAGATCAACAGGAGGGA 295

QY 51 GlnArgPhe-----SerGluIleLeuLeuLysTyr 60
Db 296 GACACTTTCTTACATCAAAACCTCCACCACGTCGCGCACACAGAGATT----- 343

QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeu-----ArgGlyPhe 77
Db 344 AACTTCAAGTGGGAGGAGTAGAGACAGACAGTGGATGGGAGAGCGCAAAAGCTGG 403

QY 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
Db 404 GCAGAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 448

QY 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAsp 117
Db 449 -----GGAGAGAGGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 472

QY 118 IleLys-----PheAspAspGluArgProHisLysArgArgLysSerArg 132
Db 473 GCAAAAGAGACAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 532

QY 133 SerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsn 152
Db 533 GACTACAGCACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592

```





```

Percent Similarity: 49.49%      Conservative: 26
Best Local Similarity: 23.23%    Mismatches: 45
Query Match: 11.05%             Indels: 5
DB: 9                            Gaps: 2

US-09-868-352-23 (1-192) x AI404883 (1-471)

QY 91 SerArgLeuGluAspTyrIle---LysGluTyrCysAsnPheGlyCysAlaTyrPheVal 109
   ||| :: ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 34 TCTGTGATAAATGATTATTTTGTAAAGCGTATTGTAATTTTTCACATATTTTTCCTCC 93

QY 110 LeuGluAsnProAsnProArgAspIleLysPheAspGluArgProHisLysArgArg 129
   ||| ||| :::::::::: :: :::::::::: ||||| ||| ||||| ||| ||||| |||
Db 94 CGTCCCAATGCAATAATAATGTTAACTAGAAAAAATAAATAAATAAATAAATAAATAA 153

QY 130 LysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSer 149
   ||| :: ||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 154 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 213

QY 150 AsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArgGlnGluArg 169
   ||||| ::||| ||||| ::||| ||||| ::||| ||||| ::||| ||||| ::|||
Db 214 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 261

QY 170 HistLeysGluGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuLeuPhe 188
   ||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 262 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 318

RESULT 5
BI082113
LOCUS      951 bp      mRNA      linear      EST 20-JUN-2001
DEFINITION      502877276F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5008939 5',
                mRNA sequence.
ACCESSION      BI082113
VERSION        BI082113.1 GI:14500443
KEYWORDS       EST.
SOURCE         house mouse.
ORGANISM       Mus musculus
                Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE      NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS        National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE          Unpublished (1999)
JOURNAL        Contact: Robert Strausberg, Ph.D.
COMMENT        Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: CLONTECH Laboratories, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: L14M1054 row: g column: 20
                High quality sequence stop: 420.

FEATURES
    source
    1..951
        /organism="Mus musculus"
        /strain="FVB/N-3"
        /db_xref="taxon:10090"
        /clone="IMAGE:5008939"
        /clone_lib="NCI_CGAP_Mam2"
        /tissue_type="tumor, biopsy sample"
        /dev_stage="5 months"
        /lab_host="DH10B"
        /notes="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Salt;
                Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigator
                providing samples: Gilbert Smith, NIH"
    BASE COUNT      295 a      255 c      330 g      71 t
    ORIGIN

Alignment Scores:      0.172      Length:      951
Pred. No.:

Score: 111.50      Matches: 27
Percent Similarity: 61.54%      Conservative: 21
Best Local Similarity: 34.62%      Mismatches: 19
Query Match: 11.05%      Indels: 11
DB: 13              Gaps: 3

US-09-868-352-23 (1-192) x BI082113 (1-951)

QY 116 ArgAspIleLysPheAspAsp-----GluArgProHis----- 126
   ||||| ||| ::||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 471 CGAGACCACAAAGAGGAGGACAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 530

QY 127 LysArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArg 146
   ||| ||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 531 AAAGAGCGCAGACAAACACAGACACAGACAGCGGAGGAGGAGGAGGAGGAGGAGGAG 590

QY 147 SerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 166
   ::||| ||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 591 AACTGAGCGCAGAGCCAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAG 650

QY 167 GlnGluArgHisLysGluGlnAspLysGluMetThrSerAlaLysGln 184
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 651 AAAGAAAG-----AAGGAGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 698

RESULT 6
BG534042
LOCUS      676 bp      mRNA      linear      EST 03-APR-2001
DEFINITION      602553149F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4662805 5',
                mRNA sequence.
ACCESSION      BG534042
VERSION        BG534042.1 GI:13525582
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS        National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE          Unpublished (1999)
JOURNAL        Contact: Robert Strausberg, Ph.D.
COMMENT        Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: CLONTECH Laboratories, Inc.
                cDNA Library Preparation: CLONTECH Laboratories, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: L14M1464 row: a column: 14
                High quality sequence stop: 231.

FEATURES
    source
    1..676
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4662805"
        /clone_lib="NIH_MGC_77"
        /lab_host="DH10B (T1 phage-resistant)"
        /notes="Organ: lung; Vector: pDNR-LIB (Clontech); Site:1:
                SfiI (ggcgctcgcc); Site:2: SfiI (ggccattatggcc); 5' and
                3' adaptors were used in cloning as follows: 5' adaptor
                sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence:
                5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size 1.9
                kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                by PCR. This library was enriched for full-length clones
                and was constructed by Clontech Laboratories (Palo Alto,
                CA). Note: this is a NIH_MGC Library."
    BASE COUNT      336 a      133 c      145 g      62 t
    ORIGIN

Alignment Scores:      0.134      Length:      676
Pred. No.:

```

```

5', mRNA sequence.
B0887742
VERSION B0887742.1 GI:22279756
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Contact: cgapbs-r@mail.nih.gov
Tissue procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM13744 row: i column: 12
High quality sequence stop: 656.
FEATURES
    source
        location/Qualifiers
            1..951
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="IMAGE:6314963"
                /clone_lib="NIH-MGC_129"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: olfactory epithelium; Vector:
                pCMV-SPORT6.1.ccdB; Site.1: EcorV; Site.2: NotI; Cloned
                unidirectionally. Primer: Oligo dr. Average insert size
                2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                is a NIH-MGC Library."
BASE COUNT 284 a 231 c 293 g 141 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 0.107 Length: 951
Score: 113.50 Matches: 58
Percent Similarity: 35.97% Conservative: 33
Best Local Similarity: 22.92% Mismatches: 71
Query Match: 11.25% Indels: 91
DB: 14 Gaps: 12
US-09-868-352-23 (1-192) x B0887742 (1-951)
Qy 14 LysTyrProGlyProGln-----PheIleHisPheGluAsnIleValLysSerAspAsp 31
Db 6 GAATTCGGGATCAAGAGTGGCTTTGTGGAGTTTGATGATCGTGGATGCAGACCAT 65
Qy 32 IleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrVal----- 48
Db 66 GCTGTTTATGAATGACGCGCAAGACCTGTGTGCGAGCGAGTCATGTTGAGCAGCCT 125
Qy 49 -----PheGlyClnArgPhe 53
Db 126 CGAGCGCGCGGGATGGCAGCTACCGCTCGGACCGCAGTGGATGGTATCGAAGA 185
Qy 54 Ser-----GluIleLeuLeuLysTyrAspPheIleValGlyAsp 66
Db 186 AGTGGCGGAGATAATACGGTCCCTCTACTCGCAGAGATACAGACTTATTGTGGAGAT 245
Qy 67 -----TrpGlyAsn-----GluGlnLeuArgLeuArgGly----- 76
Db 246 TTGTCAAGTCGATGCGTGGCAGGACCTAAAGGATTACATGCGTCAGGACGAGAGTG 305
Qy 77 PheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyr 96
Db 306 ACGTATGAGATGCTCACAAGGACGCAAAAC-----GAAGA 344
Qy 97 IleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu----- 111

```

---

```

Db 345 GTGATTGAATTGTCTTACTCTGATATGAAAGAGCTTTGAAAAGCTGGACGGAAC 404
Qy 112 AsnProAsnProAsnGAspIleLysPheAspAspGluArgPro----- 125
Db 405 GAAGTCAACAGCGCAGAAAATCAGATTAGTTGAAGACAAGCCAGGTTCTAGCGGCGCCGG 464
Qy 126 -----HisLysArg-----ArgLys 130
Db 465 TCTTACTCCAGAGCGCGAGTCACTCAAGGTCTCGCTCTCGAAGCAGACATTCGCGGAG 524
Qy 131 SerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsn 150
Db 525 AGCAGAAGCGGAGTGGTACAGTAAAGACGAGCCATTCGAAGAGCGCGTCTCGATCCAGG 584
Qy 151 AlaAsnAlaHisPheThrSerLysLysArg----- 160
Db 585 TCAGGTTCCCACTCCCGCAGCAAGAGCGCGAGTCCAGTCAGAGCGCGCAGCAGTAAG 644
Qy 161 -----LysAspThrLysArgArgGlnGlnLysLysLys 172
Db 645 AAGGAGAAAGCGCGAGCCCGCAGCAACACACACAGACCGCGCAGTCGC-----AGC 695
Qy 173 GluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Db 696 CGCAGCCCTGACAGAGCGCGCAGTAAGAGTAAGACCAAC 734
RESULT 4
LOCUS A1404883 471 bp mRNA linear EST 23-APR-2001
DEFINITION GH24733.5prime GH Drosophila melanogaster head pOT2 Drosophila
melanogaster cDNA clone GH24733 5, mRNA sequence.
ACCESSION A1404883
VERSION A1404883.1 GI:4247970
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 471)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: GH24733.3prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: GH.247 row: C column: 9
High quality sequence stop: 322
POLYA=No.
FEATURES
    source
        1..471
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="GH24733"
            /clone_lib="GH Drosophila melanogaster head pOT2"
            /sex="male and female"
            /dev_stage="adult"
            /lab_host="DH5 - alpha"
            /note="Organ: head; Vector: pOT2; Site.1: EcoRI; Site.2:
            XhoI; Sized fractionated cDNAs were directly ligated into
            pOT2. Plasmid cDNA library."
BASE COUNT 289 a 36 c 27 g 119 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0809 Length: 471
Score: 111.50 Matches: 23

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 10:34:18 : Search time 1942 Seconds  
(without alignments)  
1601.202 Million cell updates/sec

Title: US-09-868-352-23  
Perfect score: 1009  
Sequence: 1 MRKEVTPPEMLNKNYPGPQF.....EEODKEMTSAKQHLFLVRKN 192

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09868352/runat\_14022003\_153516\_4230/app\_query.fasta\_1.391  
-DB=EST -QPMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pcpt -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868352 @CGN\_1\_1\_1349 @runat\_14022003\_153516\_4230 -NCPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estnu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_htc: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_htc: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estom: \*  
17: gb\_gss: \*  
18: em\_gss\_hum: \*  
19: em\_gss\_inv: \*  
20: em\_gss\_pin: \*  
21: em\_gss\_vrt: \*  
22: em\_gss\_fun: \*  
23: em\_gss\_mam: \*  
24: em\_gss\_mus: \*  
25: em\_gss\_other: \*  
26: em\_gss\_pro: \*  
27: em\_gss\_rod: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	200.5	19.9	4933	17	BH771017	BH771017 LLMGtag73
2	113.5	11.2	686	12	BE879112	BE879112 601485839
3	113.5	11.2	951	14	BQ887742	BQ887742 AGENCOURT
4	111.5	11.1	471	9	AI404883	AI404883 GH24733.5
5	111.5	11.1	951	13	BI082113	BI082113 60287276
6	111	11.0	676	12	BG534042	BG534042 602553149
7	111	11.0	1089	12	BG850731	BG850731 102402981
c	110.5	11.0	551	10	AW553719	AW553719 L0230612
9	110.5	11.0	999	12	BE899526	BE899526 601681586
10	109.5	10.9	505	9	AL640539	AL640539 AL640539
11	109.5	10.9	1005	12	BG259084	BG259084 602379137
12	109.5	10.9	1383	12	BG288874	BG288874 602383816
13	109.5	10.9	1487	12	BE874969	BE874969 601487528
14	109	10.8	814	12	BF684615	BF684615 602140956
15	109	10.8	1256	12	BG121867	BG121867 602350882
16	107.5	10.7	471	14	BM880093	BM880093 ku04h04.y
17	107.5	10.7	1527	12	BF672136	BF672136 602152592
18	107	10.6	1091	13	BC971957	BC971957 602841742
19	106.5	10.6	511	9	AA838525	AA838525 oe90h12.s
20	106.5	10.6	966	10	BB619189	BB619189 BB619189
21	106.5	10.6	1932	12	BG114689	BG114689 602315330
22	106	10.5	831	13	BG922949	BG922949 602823807
23	106	10.5	878	13	BI552711	BI552711 603193962
24	106	10.5	1148	12	BG490668	BG490668 602520145
25	106	10.5	1260	12	BF796631	BF796631 602239829
26	106	10.5	1572	12	BG285821	BG285821 602380913
27	106	10.5	1728	12	BF578008	BF578008 602031842
28	106	10.5	1858	12	BG288695	BG288695 602385507
29	105.5	10.5	731	12	BG830845	BG830845 602767576
30	105	10.4	613	12	BG086260	BG086260 h3123504
31	105	10.4	615	14	BQ569343	BQ569343 g1126h02
32	105	10.4	820	13	BJ111917	BJ111917 BJ111917
33	105	10.4	978	12	BE914647	BE914647 601665155
34	105	10.4	1122	12	BG468632	BG468632 602510232
35	105	10.4	1173	12	BG519043	BG519043 602577966
36	105	10.4	1267	12	BG114247	BG114247 602285812
37	105	10.4	388	12	BE839084	BE839084 RC6-EN013
c	104.5	10.4	622	12	BG885312	BG885312 dad40c05
39	104.5	10.4	1166	11	AK007729	AK007729 Mus muscu
40	104.5	10.4	1787	12	BG256691	BG256691 602370880
41	104.5	10.4	882	12	BF213994	BF213994 601847482
42	104	10.3	911	12	BG478911	BG478911 602525126
43	104	10.3	929	12	BG529461	BG529461 602557839
44	104	10.3	1197	13	BI733626	BI733626 603354814
45	104	10.3				

ALIGNMENTS

RESULT 1  
BH771017

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

BH771017 4933 bp DNA linear GSS 01-MAY-2002  
LLMGtag739 MG1363 Random Sequence Tag Library Lactococcus lactis  
subsp. cremoris genomic, DNA sequence.

BH771017 GI:20373974

GSS.

Lactococcus lactis subsp. cremoris.

Lactococcus lactis subsp. cremoris

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Lactococcus.

1 (bases 1 to 4933)

Bolotin.A., Ehrlich,S.D. and Sorokin,A.

Studies of genomes of dairy bacteria Lactococcus lactis

THIS PAGE BLANK (USPTO)

Search completed: February 25, 2003, 11:35:08  
Job time : 3159 secs



ORGANISM  
Rattus norvegicus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS  
1 (bases 1 to 168111)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Fallis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegad,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,  
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

REFERENCE  
AUTHORS  
2 (bases 1 to 168111)  
Worley,K.C.  
Direct Submission  
Submitted (17-Sep-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
3 (bases 1 to 168111)  
Worley,K.C.  
Direct Submission  
Submitted (11-Jul-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 10, 2002 this sequence version replaced gi:17943731.

COMMENT  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GEIL  
Center clone name: CH230-59H10  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 107959 bases at least Q40



REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 132934)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Avele,M., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P., Buhatz,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delantey,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragune,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R.

Direct Submission  
 2 (bases 1 to 132934)  
 Worley,K.C.

Direct Submission  
 Submitted (03-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 132934)  
 Worley,K.C.

Direct Submission  
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jul 14, 2002 this sequence version replaced gi:20799942.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GVM1  
 Center clone name: CH230-357A23  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 105131 bases at least Q40  
 Consensus quality: 108098 bases at least Q30

Consensus quality: 109837 bases at least Q20

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 40 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 1186: contig of 1186 bp in length  
 1187  
 1286: gap of unknown length  
 1287  
 2398: contig of 1112 bp in length  
 2399  
 2498: gap of unknown length  
 2499  
 3572: contig of 1074 bp in length  
 3573  
 3672: gap of unknown length  
 3673  
 4761: contig of 1089 bp in length  
 4762  
 4861: gap of unknown length  
 4862  
 6408: contig of 1547 bp in length  
 6409  
 6508: gap of unknown length  
 6509  
 8289: contig of 1781 bp in length  
 8290  
 8389: gap of unknown length  
 8390  
 9557: contig of 1168 bp in length  
 9558  
 9657: gap of unknown length  
 9658  
 10777: contig of 1120 bp in length  
 10778  
 10877: gap of unknown length  
 10878  
 12603: contig of 1726 bp in length  
 12604  
 12703: gap of unknown length  
 12704  
 14881: contig of 2178 bp in length  
 14882  
 14981: gap of unknown length  
 14982  
 16572: contig of 1591 bp in length  
 16573  
 16672: gap of unknown length  
 16673  
 18014: contig of 1342 bp in length  
 18015  
 18114: gap of unknown length  
 18115  
 20339: contig of 2225 bp in length  
 20340  
 20439: gap of unknown length  
 20440  
 22021: contig of 1582 bp in length  
 22022  
 22121: gap of unknown length  
 22122  
 24163: contig of 2042 bp in length  
 24164  
 24263: gap of unknown length  
 24264  
 26505: contig of 2242 bp in length  
 26506  
 26605: gap of unknown length  
 26606  
 28894: contig of 2289 bp in length ;  
 28895  
 31231: contig of 2237 bp in length  
 31232  
 31331: gap of unknown length  
 31332  
 33251: contig of 1920 bp in length  
 33252  
 33351: gap of unknown length  
 33352  
 34661: contig of 1310 bp in length  
 34662  
 34761: gap of unknown length  
 34762  
 36350: contig of 1589 bp in length  
 36351  
 36450: gap of unknown length  
 36451  
 38161: contig of 1711 bp in length  
 38162  
 38261: gap of unknown length  
 38262  
 41430: contig of 3169 bp in length  
 41431  
 41530: gap of unknown length  
 41531  
 43197: contig of 1667 bp in length  
 43198  
 43297: gap of unknown length  
 43298  
 45621: contig of 2324 bp in length  
 45622  
 45721: gap of unknown length  
 45722  
 48922: contig of 3201 bp in length  
 48923  
 49022: gap of unknown length  
 49023  
 51372: contig of 2350 bp in length  
 51373  
 51472: gap of unknown length  
 51473  
 54730: contig of 3258 bp in length  
 54731  
 54830: gap of unknown length  
 54831  
 57633: contig of 2803 bp in length  
 57634  
 57733: gap of unknown length  
 57734  
 60622: contig of 2889 bp in length  
 60623  
 60722: gap of unknown length  
 60723  
 63956: contig of 3234 bp in length





Consensus quality: 226193 bases at least Q40  
Consensus quality: 230998 bases at least Q30  
Consensus quality: 230706 bases at least Q20  
Insert size: 235755; sum-of-contigs  
Insert size: 210251; 31.5% error; agarose-fp  
Quality coverage: 5.36x in Q20 bases; sum-of-contigs Quality  
coverage: 7.60x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 4734: contig of 4734 bp in length  
\* 4735 4834: gap of 100 bp  
\* 4835 12863: contig of 8029 bp in length  
\* 12864 12963: gap of 100 bp  
\* 12964 15044: contig of 2081 bp in length  
\* 15045 15144: gap of 100 bp  
\* 15145 20900: contig of 5756 bp in length  
\* 20901 21000: gap of 100 bp  
\* 21001 23451: contig of 2451 bp in length  
\* 23452 23551: gap of 100 bp  
\* 23552 25698: contig of 2147 bp in length  
\* 25699 25798: gap of 100 bp  
\* 25799 28077: contig of 2279 bp in length  
\* 28078 28177: gap of 100 bp  
\* 28178 31157: contig of 2980 bp in length  
\* 31158 31257: gap of 100 bp  
\* 31258 36878: contig of 5621 bp in length  
\* 36879 36978: gap of 100 bp  
\* 36979 41373: contig of 4395 bp in length  
\* 41374 41473: gap of 100 bp  
\* 41474 52993: contig of 11526 bp in length  
\* 53000 53099: gap of 100 bp  
\* 53100 55405: contig of 2306 bp in length  
\* 55406 55505: gap of 100 bp  
\* 55506 58338: contig of 2833 bp in length  
\* 58339 58438: gap of 100 bp  
\* 58439 61539: contig of 3101 bp in length  
\* 61540 61639: gap of 100 bp  
\* 61640 64042: contig of 2403 bp in length  
\* 64043 64142: gap of 100 bp  
\* 64143 66210: contig of 2068 bp in length  
\* 66211 66310: gap of 100 bp  
\* 66311 69150: contig of 2840 bp in length  
\* 69151 69250: gap of 100 bp  
\* 69251 71427: contig of 2177 bp in length  
\* 71428 71527: gap of 100 bp  
\* 71528 142636: contig of 71109 bp in length  
\* 142637 142736: gap of 100 bp  
\* 142737 162663: contig of 19927 bp in length  
\* 162664 162763: gap of 100 bp  
\* 162764 177276: contig of 14513 bp in length  
\* 177277 177376: gap of 100 bp  
\* 177377 237855: contig of 60479 bp in length.

FEATURES

source

1. 237855  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-192D3"  
/clone\_lib="RPC1-11.1"  
1. 4734  
/note="assembly fragment:03221"  
fragment\_chain:1  
4835. 12863  
/note="assembly fragment:03852"  
fragment\_chain:1  
12964. 15044

misc\_feature /note="assembly\_fragment:00456"  
15145. 20900  
/note="assembly\_fragment:00478"  
21001. 23451  
/note="assembly\_fragment:00712"  
23552. 25698  
/note="assembly\_fragment:01001"  
25799. 28077  
/note="assembly\_fragment:01623"  
28178. 31157  
/note="assembly\_fragment:01683"  
31258. 36878  
/note="assembly\_fragment:02162"  
36979. 41373  
/note="assembly\_fragment:02751"  
41474. 52999  
/note="assembly\_fragment:03295"  
53100. 55405  
/note="assembly\_fragment:03484"  
55506. 58338  
/note="assembly\_fragment:03568"  
58439. 61539  
/note="assembly\_fragment:03629"  
61640. 64042  
/note="assembly\_fragment:03661"  
64143. 66210  
/note="assembly\_fragment:03961"  
66311. 69150  
/note="assembly\_fragment:04258"  
69251. 71427  
/note="assembly\_fragment:04367"  
71528. 142636  
/note="assembly\_fragment:02579"  
fragment\_chain:2  
142737. 162663  
/note="assembly\_fragment:01298"  
fragment\_chain:2  
162764. 177276  
/note="assembly\_fragment:03624"  
fragment\_chain:2  
177377. 237855  
/note="assembly\_fragment:02137"  
fragment\_chain:2  
clone\_end:SP6  
vector\_side:right  
BASE COUNT 73514 a 42740 c 42430 g 77064 t 2107 others  
ORIGIN

Alignment Scores:  
Pred. No.: 335 Length: 237855  
Score: 106.50 Matches: 38  
Percent Similarity: 43.79% Conservative: 36  
Best Local Similarity: 22.49% Mismatches: 78  
Query Match: 10.56% Indels: 17  
DB: 2 Gaps: 4

US-09-868-352-23 (1-192) x AL391558 (1-237855)

Qy 33 GluPheGlnLeuValIleAsnGluLysSer-----AlaPheAsp-----ValThr 47  
Db 32976 GAAATCAGCTGCTCCAGCCTCAAAAGCTTTTGCAGCAATTTGACACCAACATCACT 33035  
Qy 48 ValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIleValGlyAspTrp 67  
Db 33036 GTTAAACAACGAAACCTTAATGATCAAAATTCAAATATAGATCATCAATTTCTGGAATA 33095  
Qy 68 GlyAsnGluGlnLeuArgLeuArgLeuGlyPheTyrIleLysAspAlaSerThrIleArgLysAsn 87  
Db 33096 TCAGCGACAGAACTGAGTATCAAGACTATCTACCGCAT----- 33134  
Qy 88 SerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyr 107  
Db 33135 CAAGACTGGAGATCAATAATAGTTATATATTTGAAGAATACTATGGAATAAATATAC 33194

CDS

```

/codon_start=1
/product="hypothetical protein"
/protein_id="AAL93606.1"
/db_xref="GI:1970314"
/translation="MKLIILVIFLKLISINGFENIDGNNELIET
NKDFKINGSSICFSGCGGFWNVFSSGLNGSEIFVKSEEDY
KIDTLHVPNOKLRNKTPTVFQREYVYSGCFNVLKNCIDKID
KIISPLNHSHTPSPIHFTFTFLVVFQFNTSNLSFLLIIN
PAIFNCAJFFSTTEILGLSAYLVENSQFYGGEARNPKSNVGV
FSNKLIFSLIDIQQCSFLVAPKFIKSIWIISIPISSEIIFIV
DSNFLFYLLIIVYHPCKSNLIVLRNLKSYVDGDELSINQNSQ
NK"
complement(join(17079..19519,19709..20690)))
cds

```

```

/coords/Star_C-1
/product="hypothetical_protein"
/protein_id="AAL93607.1"
/db_xref="GI:19703145"
/translation="MKTHKKKSLGGFLSHSSSSPNKSLFTEEVHEQQOQQOQHNN
NNNNNHQRPQSTTSVYSDASASIEPSSYLSKTSSSSLSPESSPLHQNQNG
NIQLNSSGSLNNEQQOQQQSTQTSNPNTSPIMGKKKPSRLAQAFKRVKNNGNK
LKEIEELNKTGLNQYTHSNLPFVNVESDDIDNGSGGTTSTGNIISHKSPSSSS
SSSSCKHQKRSRPIEIPISDTEPDYDIPRIIKMSVEYLFKCLLVGPIFRSANAME
LQRLSQLFERGGIDISTYDHPICIGGLKLKLYFREKPIPFYDGLHKRIYSVLNEDES
NIKVSLLSGELGKGQGLILRYLFLNLAVNSINFMFMYQNLCAIPASLIQFDEL
SCYDVEIRLTENVHISFVGVPVVDNNSVSSSLAHPILIGGKSETNYSNLNLTVI
DHGIDINNNSNNNSSSSSPYKSPKSPSKLPPNRRNSISPKLSSTSPFKSI
NLSNSTMIHSDVQOQQOQQOQDEQDEQDEQDEQDEQNSNSTSSSTIT
APRKSSTQYKLNRLINTCRPSSWNNNRILQOQHHHHHHHQOQQHQOQQOSSSES
NLSSTSPQRLNTRVLESYEGKWNNIYISLKEESKLPPKSLNMQITVNSNI
IGKNGRSDCTQSDDEELIKSPITSTVNEILKETNDNDNDQNTQNNSSNNI
PKTITITNTNTTNNIPVAKSQIITSTSAKVPTPTPAPMOTSSFLSTQTNPS
SSSSPSSSTSSNNISDNKNTSNVSNYNNFQAPNRTVSSPNVRNFSVPTT
TTTTIGNSFEVSPREIGKRTTSPISFPKQSIKCPKNTTSLSSSSNNISKSTNPT
TPTPAPLITLTSVNSTSSSKLQKALPKTFTTQTSPLPILLPOPLTSTSTSTST
TSTSTSTSTSTSTPTTSPFSNFSFTSDNNSNNNNNNNNNNNNNNNNNNNN
NNNNNTPTIVSPSSSTSPISPTSTIKGHSNLSLSLQTPSSSSSSSLGRLTAPG
ROLNHTNNTSLSKFNQIDTSGINQGWKGGVVYQKQDPVKKSLAIPEK"
complement(join(22671, 22910, 23013, 23359, 23480, 23555))

```

```

/coodon_start=1
/product="hypothetical protein"
/protein_id="AAL93608.1"
/db_xref="GI:19703146"
/translation="MKVILPLVSLIVLISAVCTIIIGFTYMQVQVGVSGSTVTTY
YKYNKVSLSFNGDEETSDYDDGFESAKPKFYKAALSDMLAFVNLFSACYM
GMAAFKGISIPANTQKLNLSNGATFVLIATFVVLGPKASRDECDYAKDSDYTFG
SQVDMCGGDIYKKEIGSHYDIEDGSKLSWGPNTAWIAIIVVGLGIATFNFINFVYCGPVG

```

[illegible]

Alignment Scores:		
Pred. No.:	19.1	Length:
Score:	106.50	Matches:
Percent Similarity:	43.14%	Conservative:
Best Local Similarity:	18.14%	Mismatches:
Query Match:	10.56%	Indels:
DB:	2	Gaps:
		25117

US-09-868-352-23 (1-192) x AC116102 (1-25117)

Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20

[illegible]

RESULT 37					
AL391558					
LOCUS	AL391558	237855 bp	DNA	linear	HTG 10-JUL-2001
DEFINITION	Homo sapiens chromosome 9 clone RP11-192D3, *** SEQUENCING IN PROGRESS *** 22 unordered pieces.				

AL391538  
AL391558.6  
GI:10716369  
HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
human

SOURCE  
HUMANI.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 237855)

AUTHORS	BURTON, J.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT	On Oct 7, 2000 this sequence version replaced a1:10277996.

```

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bA132D3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752;
Chemistry: Dye-terminator Big Dye; 1

```



Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehocsky, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talames, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 27, 2002 this sequence version replaced gi:18482207.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L21052  
Center clone name: 199\_E\_23  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 201345 bases at least Q40  
Consensus quality: 209044 bases at least Q30  
Consensus quality: 212024 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 213377; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 54 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 12: contig of 12 bp in length  
\* 13 112: gap of 100 bp  
\* 113 767: contig of 655 bp in length  
\* 768 867: gap of 100 bp  
\* 868 1486: contig of 619 bp in length  
\* 1487 1586: gap of 100 bp  
\* 1587 2581: contig of 995 bp in length  
\* 2582 2681: gap of 100 bp  
\* 2682 3615: contig of 934 bp in length  
\* 3616 3715: gap of 100 bp  
\* 3716 4456: contig of 741 bp in length  
\* 4457 4556: gap of 100 bp  
\* 4557 5548: contig of 992 bp in length  
\* 5549 5648: gap of 100 bp  
\* 5649 6452: contig of 804 bp in length  
\* 6453 6552: gap of 100 bp  
\* 6553 7198: contig of 646 bp in length  
\* 7199 7298: gap of 100 bp  
\* 7299 8131: contig of 833 bp in length  
\* 8132 8231: gap of 100 bp  
\* 8232 9099: contig of 868 bp in length

9100 9199: gap of 100 bp  
9200 10023: contig of 824 bp in length  
10024 10123: gap of 100 bp  
10124 11265: contig of 1142 bp in length  
11266 11365: gap of 100 bp  
11366 11997: contig of 632 bp in length  
11998 12097: gap of 100 bp  
12098 13106: contig of 909 bp in length  
13107 13106: gap of 100 bp  
13107 14298: contig of 1192 bp in length  
14299 14398: gap of 100 bp  
14399 15431: contig of 1033 bp in length  
15432 15531: gap of 100 bp  
15532 16342: contig of 811 bp in length  
16343 16442: gap of 100 bp  
16443 17510: contig of 1068 bp in length  
17511 17610: gap of 100 bp  
17611 18866: contig of 1256 bp in length  
18867 18966: gap of 100 bp  
18967 19993: contig of 1027 bp in length  
19994 20093: gap of 100 bp  
20094 20938: contig of 845 bp in length  
20939 21038: gap of 100 bp  
21039 22307: contig of 1269 bp in length  
22308 22407: gap of 100 bp  
22408 23319: contig of 912 bp in length  
23320 23419: gap of 100 bp  
23420 24672: contig of 1253 bp in length  
24673 24772: gap of 100 bp  
24773 25600: contig of 828 bp in length  
25601 25700: gap of 100 bp  
25701 27029: contig of 1329 bp in length  
27030 27129: gap of 100 bp  
27130 28673: contig of 1544 bp in length  
28674 28773: gap of 100 bp  
28774 29795: contig of 1022 bp in length  
29796 29895: gap of 100 bp  
29896 31567: contig of 1672 bp in length  
31568 31667: gap of 100 bp  
31668 33129: contig of 1462 bp in length  
33130 33229: gap of 100 bp  
33230 34638: contig of 1409 bp in length  
34639 34738: gap of 100 bp  
34739 36991: contig of 2253 bp in length  
36992 37091: gap of 100 bp  
37092 39530: contig of 2439 bp in length  
39531 39630: gap of 100 bp  
39631 41804: contig of 2174 bp in length  
41805 41904: gap of 100 bp  
41905 44637: contig of 2733 bp in length  
44638 44737: gap of 100 bp  
44738 47882: contig of 3145 bp in length  
47883 47982: gap of 100 bp  
47983 51826: contig of 3844 bp in length  
51827 51926: gap of 100 bp  
51927 56275: contig of 4349 bp in length  
56276 56375: gap of 100 bp  
56376 61809: contig of 5434 bp in length  
61810 61909: gap of 100 bp  
61910 68212: contig of 6303 bp in length  
68213 68312: gap of 100 bp  
68313 74425: contig of 6113 bp in length  
74426 74525: gap of 100 bp  
74526 79707: contig of 5182 bp in length  
79708 79807: gap of 100 bp  
79808 86515: contig of 6708 bp in length  
86516 86615: gap of 100 bp  
86616 93246: contig of 6631 bp in length  
93247 93346: gap of 100 bp  
93347 101225: contig of 7879 bp in length  
101226 101325: gap of 100 bp  
101326 112385: contig of 11060 bp in length  
112386 112485: gap of 100 bp

TITLE  
JOURNAL

## COMMENT







## RESULT 32

```

AC083856
LOCUS
DEFINITION Mus musculus chromosome 12 clone RP23-366H13 strain C57BL6/J,
            215455 bp DNA linear HTG 04-OCT-2000
            WORKING DRAFT SEQUENCE, 16 unordered pieces.
ACCESSION AC083856
VERSION AC083856.1 GI:10567854
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 215455)
AUTHORS Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
          Bouffard,G.G., Dietrich,N.L., Gupta,J., Ho,S.-L., Idol,J.,
          Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
          Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Snyder,B.,
          Stantripop,S., Thomas,J.W., Thomas,P.J., Tlionson,E.E.,
          Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A., Wetherby,K.D. and
          Green,E.D.
TITLE NISC Mouse Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 215455)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2000) NIH Intramural Sequencing Center, 8717
          Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT ----- Genome Center
          Center: NIH Intramural Sequencing Center
          Center code: NISC
          Web site: http://www.nisc.nih.gov
          Contact: nisc.mouse@phgrl.nih.gov
          ----- Project Information
          Center project name: sh
          Center clone name: 366H13
          ----- Summary Statistics
          Sequencing vector: plasmid; n/a; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.990319
          Consensus quality: 207240 bases at least Q40
          Consensus quality: 209922 bases at least Q30
          Consensus quality: 211032 bases at least Q20
          Insert size: 214000; agarose-fp
          Insert size: 229000; pulse-field-gel
          Insert size: 213955; sum-of-contigs
          Quality coverage: 5.63x in Q20 bases; agarose-fp
          Quality coverage: 5.26x in Q20 bases; pulse-field-gel
          Quality coverage: 5.63x in Q20 bases; sum-of-contigs
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 16 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          * 1
          * 5772: contig of 5772 bp in length
          * 5773
          * 5872: gap of unknown length
          * 5873
          * 10133: contig of 4261 bp in length
          * 10134
          * 10233: gap of unknown length
          * 10234
          * 15586: contig of 5353 bp in length
          * 15587
          * 15686: gap of unknown length
          * 15687
          * 20246: contig of 4560 bp in length
          * 20247
          * 20346: gap of unknown length
          * 20347
          * 26852: contig of 6506 bp in length
          * 26853
          * 26952: gap of unknown length
          * 26953
          * 36105: contig of 9153 bp in length
          * 36106
          * 36205: gap of unknown length
          * 36206
          * 44497: contig of 8292 bp in length
          * 44498
          * 44597: gap of unknown length
          * 44598
          * 53137: contig of 8540 bp in length
          * 53138
          * 53237: gap of unknown length
          *
          * 53238
          * 62667: contig of 9430 bp in length
          * 62668
          * 62767: gap of unknown length
          * 71276: contig of 8509 bp in length
          * 71277
          * 71376: gap of unknown length
          * 81084: contig of 9708 bp in length
          * 81085
          * 81184: gap of unknown length
          * 81185
          * 95747: contig of 14563 bp in length
          * 95847: gap of unknown length
          * 95848
          * 112911: contig of 17064 bp in length
          * 112912
          * 113011: gap of unknown length
          * 113012
          * 132279: contig of 19268 bp in length
          * 132280
          * 132379: gap of unknown length
          * 132380
          * 156697: contig of 27318 bp in length
          * 156698
          * 156998
          * 215455: contig of 55658 bp in length.
          *
          * Location/Qualifiers
          * 1..215455
          * /organism="Mus musculus"
          * /strain="C57BL6/J"
          * /db_xref="taxon:10090"
          * /chromosome="12"
          * /clone_lib="RPCI mouse BAC library 23"
          * /clone="RP23-366H13"
          * 1..5772
          * /note="assembly_fragment"
          * 5873..10133
          * /note="assembly_fragment"
          * /note="assembly_fragment"
          * /note="end:SP6"
          * /vector_side:right"
          * 10234..15586
          * /note="assembly_fragment"
          * 15687..20246
          * /note="assembly_fragment"
          * /note="assembly_fragment"
          * /note="end:T7"
          * /vector_side:right"
          * 20347..26852
          * /note="assembly_fragment"
          * 26953..36105
          * /note="assembly_fragment"
          * /note="assembly_fragment"
          * 36206..44497
          * /note="assembly_fragment"
          * 44598..53137
          * /note="assembly_fragment"
          * 53238..62667
          * /note="assembly_fragment"
          * /note="assembly_fragment"
          * 62768..71276
          * /note="assembly_fragment"
          * /note="assembly_fragment"
          * 71377..81084
          * /note="assembly_fragment"
          * 81185..95747
          * /note="assembly_fragment"
          * 95848..112911
          * /note="assembly_fragment"
          * 113012..132279
          * /note="assembly_fragment"
          * /note="assembly_fragment"
          * 132380..156697
          * /note="assembly_fragment"
          * /note="assembly_fragment"
          * 156998..215455
          * /note="assembly_fragment"
          * /note="assembly_fragment"
          *
          * BASE COUNT 68411 a 40237 c 39791 g 65506 t 1510 others
          * ORIGIN
          *
          * Alignment Scores:
          * Pred. No.: 213 Length: 215455
          * Score: 108.00 Matches: 20
          * Percent Similarity: 60.87% Conservative: 22
          * Best Local Similarity: 28.99% Mismatches: 27
          * Query Match: 10.70% Indels: 0
          * DB: 2 Gaps: 0
          *
          * US-09-868-352-23 (1-192) x AC083856 (1-215455)
          * QY 116 ArgAspileLysPheAspAspGluArgProHisLysArgLysSerArgSerLysSer 135

```











```
REFERENCE 1
AUTHORS Berka,R. and Clausen,I.G.
TITLE Methods for monitoring multiple gene expression
JOURNAL Patent: WO 0229113-A 3145 11-APR-2002;
Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
FEATURES
source 1. .273
/organism="Bacillus licheniformis"
/db_xref="taxon:1402"
BASE COUNT 89 a 48 c 61 g 75 t
ORIGIN
Alignment Scores:
Pred. No.: 5.84e-13 Length: 273
Score: 222.50 Matches: 42
Percent Similarity: 74.68% Conservative: 17
Best Local Similarity: 53.16% Mismatches: 19
Query Match: 22.05% Indels: 1
DB: 6 Gaps: 1
US-09-868-352-23 (1-192) x AX434730 (1-273)
Qy 34 PheGlnLeuValIleAsnGluLysSerAlaPheValThrValPheGlyGlnArgPhe 53
Db 26 TTTGAAGTCGTCAAAGATGTGAAAGACGGTTTAAACGAAGAGCGTTTAAAGCCAGATAC 85
Qy 54 SerGluIleLeuLeuLysTyAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArg 73
Db 86 TCCGACATTTTAATAATACGATACATCTCGCGGATGCGGATGCGGATGCGGATGCGGATGCGG 145
Qy 74 LeuArgGlyPheTyLysAspAlaSer---ThrIleArgLysAsnSerArgIleSerArg 92
Db 146 CTGAAGGCTTTTTCAGCATCAAAATCAAAAGGCCACATTTGATACGAGATCAGCAGC 205
Qy 93 LeuGluAspTyIleLysGluTyCysAsnPheGlyCysAlaTyPheValLeuGlu 111
Db 206 CTTGATGAATATATTACGAGTACTGCAATTCGATGTCGCTACTTTGTTTAAAA 262
RESULT 26
LOCUS E26748 6145 bp DNA linear PAT 18-JUN-2001
DEFINITION Structure and function of novel biorhythm marker gene.
ACCESSION E26748
VERSION E26748.1 GI:13026326
KEYWORDS JP 1999169185-A/2.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 6145)
Mario,I., Yutaka,S., Satoru,S. and Mikio,K.
Structure and function of novel biorhythm marker gene
Patent: JP 1999169185-A 2 29-JUN-1999;
AGENCY OF IND SCIENCE & TECHNOL.,OTSUKA PHARMACEUT CO LTD
OS Rattus sp. (Wistar rat)
PN JP 1999169185-A/2
PD 29-JUN-1999
PF 12-DEC-1997 JP 1997362890
PR MARIO ISHIDA,YUTAKA SADAKANE,SATORU SUZUKI,MIKIO KIKUCHI PC
C12N15/09,C07K14/47,C12Q1/68//C12N5/10,C12P21/02,C12N15/09, PC
C12R1:91),
PC (C12N5/10,C12R1:91),(C12P21/02,C12R1:91),(C12N15/00,C12N5/00,
PC (C12N15/00,C12R1:91),(C12N5/00,C12R1:91)
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT CDS 3186..4365.
Location/Qualifiers
1. .6145
/organism="Rattus sp."
/db_xref="taxon:10118"
FEATURES
source
```

```
BASE COUNT 1849 a 1110 c 1364 g 1822 t
ORIGIN
Alignment Scores:
Pred. No.: 0.69 Length: 6145
Score: 113.50 Matches: 55
Percent Similarity: 37.16% Conservative: 42
Best Local Similarity: 21.07% Mismatches: 73
Query Match: 11.25% Indels: 91
DB: 6 Gaps: 11
US-09-868-352-23 (1-192) x E26748 (1-6145)
Qy 8 GluMetLeuAsnTyAsnLysTyProGlyProGlnPheIleHisPheGluAsnIleVal 27
Db 2788 AAAATGTTAAATGTAAAGCAGGTTCCTTCCCTCAAAGTTTAAATTTTCACACAGAGAA 2847
Qy 28 LysSerAspIleGlu----- 33
Db 2848 AGGCATCTAATTTAGAAAAACACAGATATTCAGTTTGTAGACTAAAGGAAACTTACT 2907
Qy 34 -----PheGlnLeuValIleAsnGluLysSerAlaPhe-----AspValThr 47
Db 2908 CAAATTTTGTGTCTAAATACTTTAAACCTCCACTCTGAACCTCTATGTGATTTATCT 2967
Qy 48 ValPheGlyGlnArgPheSerGluIleLeuLysTyAspPhe----- 62
Db 2968 TTCATT-----AGATTTCCTAAATTAATCTTGAATTTAGTTTCGGTTCCTAACAAAATA 3021
Qy 63 -----IleValGlyAsp 66
Db 3022 AATGCAGGAAAGTCTTAGCCCTCACTGCTAGGAATTCACACATGTTTAACTCTGCGT 3081
Qy 67 TrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyLysAspAlaSerThrIleArgLys 86
Db 3082 TATGGCAGCAACAGAATA---CTACGAACAGCTCTTTATAATCATCTGTTTATATGTCAC 3138
Qy 87 AsnSerArgIleSerArgLeuGluAsp-----TyrIleLys 98
Db 3139 TGTGTCAGGGTGGCTGCCTCAACGAGTTTATGCANTCACTTTCTTGGATGTTTCTGAAG 3198
Qy 99 GluTyrcysAsnPheGly----- 104
Db 3199 GATAATTGCACAGCAGGAGGATGTACAGAGAGTAGGCCCTTGCACCTATATGTGTATAT 3258
Qy 105 -----CysAlaTyrc----- 107
Db 3259 TCCACTTGTGCTTATTTAATTAACCTGGGATCTTTAATTGTTCTGAGCTTACGCTCAAAAGT 3318
Qy 108 -----PheValLeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgPro 125
Db 3319 GGTGTTTTCCTCCAGAGTCTGGAAAGAGCAATGAAGGAAAGCGGCTCGATCTCGTTC 3378
Qy 126 HisLysArgLysSerArgSerLysSerGlnSerLysSerLysSerGlnThrArgAsnAsn 145
Db 3379 CACACTCGCTCAAAAGTCCAGGTCT-----AGCTCAAAAGTCCCAATCTCAGAAGGAAG 3429
Qy 146 ArgSerGlnSer-----AsnAlaAsnAlaHisPheThrSerLysLysArg---LysAsp 162
Db 3430 AGATCGGAGTCAAAAGCAGGAGTAGATCCCATACAGGTCACGTTCAAGACAGAAAGAT 3489
Qy 163 ThrLysArgGlnGlnGluArgHisIleLysGluGlnAspLysGluMetThrSerAla 182
Db 3490 AGAGTAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3549
Qy 183 Lys 183
Db 3550 AGG 3552
RESULT 27
AC125344 179908 bp DNA linear ROD 17-AUG-2002
LOCUS AC125344
DEFINITION Mus musculus chromosome 12 clone RP24-332H21, complete sequence.
```





```

DEFINITION Sequence 7936 from Patent WO0229113.
ACCESSION AX439521
VERSION AX439521.1 GI:21664332
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1
AUTHORS Berka R. and Clausen, I.G.
TITLE Methods for monitoring multiple gene expression
JOURNAL Patent: WO 0229113-A 7936 11-APR-2002;
Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
FEATURES
Location/Qualifiers
source
1..285
/organism="Bacillus clausii"
/db_xref="taxon:79880"
BASE COUNT 93 a 51 c 66 g 74 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 6,92e-14 Length: 285
Score: 232.50 Matches: 46
Percent Similarity: 69.32% Conservative: 15
Best Local Similarity: 52.27% Mismatches: 22
Query Match: 23.04% Indels: 5
Gaps: 2
DB: 6

US-09-868-352-23 (1-192) x AX439521 (1-285)
Qy 26 lleVallySerAspAspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAsp 45
Db 1 ATGGTGCAGGATTTCAGGCATTGAATATGATCAAGTCCTCGAAACGAGCGCANTGGTGGAAAT 60
Qy 46 ValThrValPheGlyClnArgPheSerGlnIleLeuLysTyrAspPheIleValGly 65
Db 61 GAGGAGGACATTCAAGCGCGGTACAGCATGACTGAACAATATGATTACATCGTCGCGC 120
Qy 66 AspTrpGlyAsnGluIleLeuArgLeuArgGlyPheTyrLysAsp-----AlaSer 82
Db 121 GATTGGGGCTACACCACTCGGTTGCGCGCTTTTGTGATGACAAAAACAAAAATCA 180
Qy 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102
Db 181 ACATTTCGATAAA-----AAATTAGCACGTTACCGCGACTATCTTTACGAATACATGCAAT 234
Qy 103 PheGlyCysAlaTyrPheValLeu 110
Db 235 TTGTGTCGCGCTATTGTTGCTTG 258

RESULT 24
BSUB0017/c 217420 bp DNA linear BCT 26-NOV-1997
LOCUS Bacillus subtilis complete genome (section 17 of 21): from 3197001
DEFINITION to 3414420.
ACCESSION 299120 AL009126
VERSION 299120.1 GI:2635613
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1
AUTHORS Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Borris, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerion, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
Faulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,
Grandi, G., Guiseppli, G., Guy, B.J., Haga, K., Haeck, J., Harwood, C.R.,
Henauf, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,

```

```

Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetelle, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
Scanlan, E., Schleicher, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
Viari, A., Wembut, R., Wedler, E., Wedler, H., Weitzengger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
PUBMED 9384377
2 (bases 1 to 217420)
REFERENCE
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
FEATURES
Location/Qualifiers
source
1..217420
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(3..1277)
/gene="yulE"
complement(3..1277)
/gene="yulE"
/function="unknown"
/notes="similar to L-rhamnose isomerase"
/codon_start=1
/transl_table=11
/protein_id="CAB15107.1"
/db_xref="GI:2635614"
/transl_table=11
/translation="MTIKANYDSAKQAYEKWIDVEALRQLEQVPISTHWCQDDOTE
GFVNGELSGGIDVTGNYPGKATPEELRDLLEKSLIPGKRVNHLAIYAEITNRE
AVERDELKPOHFENWVKNLGLDFNPFLFSBKADGLTSLPDDPDIFSEFWIRH
CIACRIGYFGKELGTPCLNTWIPDGKIDIPSDRLTPKRLKESLDLRFSEISQ
HNLDSEKFLGSGESYVVGSHFYLAYALT/NHKLCLLDTHGHPPTVSNKISSML
LYTDLKALHVSFVRWDSHVVDLDELALEIVRNHAKLEIVAGLDFDASINRV
AAWTIGTRNNIKALLYALLPGLQKEGRTYTERLALMBEFTKYPFGAINDSYCE
QMGVPVKEAWLYDIKEYEQVLLKRASSPIV"
complement(1303..1617)
/gene="yulD"
complement(1303..1617)
/gene="yulD"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15108.1"
/db_xref="GI:2635615"
/translation="MKRKASIMFVHQDYKYEYKQRIHDDIMPENAEALKARGAHYISTF
LDSETRGLFAYLEIDEDEKWRKMADEVCQRWKNMAPLTKNTSDFSPVAIDLKEYV
LD"
complement(1635..3092)
/gene="yulC"
complement(1635..3092)
/gene="yulC"

```

```
/db_xref="GI:10175797"  
/translation="MAVLTFAFISGLFLYFKIAWRTPESLKRIHESKASTASG  
VELPFAGINLLVGLRGTFVDLIVCLVGFVGGFILIHGIRAYQHYVPQLKQG"  
complement(3379..4686)
```

[illegible]

	RESULT 23		
	AX439521		
	LOCUS	DNA	PAT 28-JUN-2002
		linear	
	AX439521	285 bp	



```

/db_xref="GI:16411816"
/db_xref="SPTREMBL:O8Y4V3"
/translation="MNRVLSNKRSEQKDCVFKTDIPPENIKSLNRMQGRNCCYG
VLEINGKQLRKGITAVKLDLKAAGSLPVVQVEYHPETISEMORLIWSGY"
complement(3092..3097)
/gene="lmo2327"
complement(3287..3540)
/gene="lmo2328"
complement(3287..3529)
/gene="lmo2328"
/note="similar to transcription regulator"
/codon_start=1
/transl_table=11
/protein_id="CAD00406.1"
/db_xref="GI:16411817"
/db_xref="SPTREMBL:O8Y4V2"
/translation="MRLWLKIREENGLTQODVAELADVETTYASIEQGRNPSVW
AMRIASVLDPEWTFIFDDKVRKSHFYLNKQEAR"
complement(3535..3540)
/gene="lmo2328"
3672..4161
/gene="lmo2329"
3672..3677
/gene="lmo2329"
3685..4161
/gene="lmo2329"
/note="similar to a putative repressor protein
[Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00407.1"
/db_xref="GI:16411818"
/db_xref="SPTREMBL:O8Y4V1"
/translation="MFGDILMKLRKNLQSDIAKILGVARTTYSSEQNRRMPDAE
IQIKADYFDVSDLYLHGRKNNIADTIAAHDSNASEDEKEILAYIEEKKRKHVNE
REINTDAASKGDKVEPNVEDNEFDKAAVAIRMSDAEAVKAVKSFIEFEQKNN"
4528..5243
/gene="lmo2330"
4528..4533
/gene="lmo2330"
4542..5243
/gene="lmo2330"
/note="similar to protein gp33 [Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00408.1"
/db_xref="GI:16411819"
/db_xref="SPTREMBL:O8Y4V0"
/translation="WSKYSYLLKKWFWILLIILSLFNGIWLIFLATLTFAL
IKVINERRRKTLIIIIATVLLIFSLIRTVOLNYIVENPEEWTIDNOKDKFAL
AETDEVEETEEDQADASAAVTYQEGIESFNRLISENGILIKVVPPENEYDM
LIAYVTDLKYODETIKQKNADYLGNEIQORALGTLFGGDSNRHPMVELRYEDETMA
GSAFDKTNMKLTGK"
5251..5944
/gene="lmo2331"
5251..5256
/gene="lmo2331"
5264..5944
/gene="lmo2331"
/note="weakly similar to gp32_Bacteriophage A118 protein"
/codon_start=1
/transl_table=11
/protein_id="CAD00409.1"
Alignment Scores:
Pred. No.: 9,34e-11 Length: 324050
Score: 240.50 Matches: 53
Percent Similarity: 67.23% Conservative: 27
Best Local Similarity: 44.54% Mismatches: 28
Query Match: 23.84% Indels: 11
DB: 1 Gaps: 4

US-09-868-352-23 (1-192) x AL591983 (1-324050)

```

```

QY 2 ArgLysGluVal---ThrProGluMetLeu--AsnTyrAsnLysTyrProGlyProGlnP 20
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 80946 CGAAAAAGAGTATATACACGGAACTACTCCGAGATATTTTAAATGGAACCTCAAAAA 80887

QY 20 heIleHisPhe-----GluAsnIleValLysSerAspIleGluPheGlnLeuV 37
:::  :::  :::  :::  :::  :::  :::  :::  :::  :::  |||||  |||||  |||||
Db 80886 ATGTACGGACAGGAGAACAGTACGATTCAGGATTCAGGATTTTAACTACGAGATTA 80827

QY 37 alIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleL 57
::  |||  :::  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 80826 TCACCAATTATCGCGATGCTTTTGACGAGGAGAACTTAATGAACGTTTACGGATATTC 80767

QY 57 euLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyP 77
||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
Db 80766 TTGGACGATATGATTATAGTGGCGACTGGGTTTACGATCACTCGGCTGAAAGGCT 80707

QY 77 heTyrLysAspAlaSerThrIleArgLys-----AsnSerArgIleSerArgLeuG 94
||:  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
Db 80706 TTTTCGAAGACGATAAC-----CGCAAGCGCGCATACGACAAATAATAGCACGTTAA 80653

QY 94 luAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111
:::  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 80652 AGCAATATATTTATGAATACTGCAATTTGGTTGCGGTATTTGTTATTATAA 80600

RESULT 20
AX414540
LOCUS AX414540 1234 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1531 from Patent WO0228891.
ACCESSION AX414540
VERSION AX414540.1 GI:21446997
KEYWORDS Listeria monocytogenes ATCC 19115.
ORGANISM Listeria monocytogenes ATCC 19115
SOURCE Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1
AUTHORS Glaser,P. and Kunst,P.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 1531 11-APR-2002;
Pasteur Institut (FR)
FEATURES
Source 1..1234
/organism="Listeria monocytogenes ATCC 19115"
/db_xref="taxon:176281"
BASE COUNT 402 a 195 c 288 g 349 t
ORIGIN
Alignment Scores:
Pred. No.: 9,68e-14 Length: 1234
Score: 239.50 Matches: 43
Percent Similarity: 77.38% Conservative: 22
Best Local Similarity: 51.19% Mismatches: 14
Query Match: 23.74% Indels: 5
DB: 6 Gaps: 2

US-09-868-352-23 (1-192) x AX414540 (1-1234)
QY 31 AspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 541 GATTAACTACGAGATATATACCAATTTATCGCGACGCTTTTGACGAGGAGAAAGTTAAAT 600

QY 51 GlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70
:::  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 601 GAACGGTTTACGATATCTTGGCGCATATGATTATATAGTAGCGCATGGGTTCAGCAT 660

QY 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys-----Asn 87
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 661 CAATCCGCTCAAAGCGCTTTTGAAGACGATAAC-----CGCAAAAGCCGCATACGAC 714

QY 88 SerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyr 107
:::  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 715 AACAAATTAGCACGTTAAAGAATATATATTACGATCTGCAATTTTGGTTGCGGTAT 774

```

```

||||| ||||||| ||||| ||||| |||||
Db 241 TGTCTTATTGTAGTCAGACGATTGTCACCAATGAATTTATTGAAGAAATAGATGAT 300
Qy 123 Glu 123
Db 301 AAA 303

RESULT 19
AL591983/c AL591983 324050 bp DNA linear BCT 06-JUN-2002
LOCUS Listeria monocytogenes strain EGD, complete genome segment 11/12.
DEFINITION AL591983 AL591824
ACCESSION AL591983.1 GI:16411809
VERSION AL591983.1 GI:16411809
KEYWORDS
SOURCE Listeria monocytogenes.
ORGANISM Listeria monocytogenes
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1
AUTHORS Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A.,
Baquero,F., Berche,P., Bloeker,H., Brandt,P., Chakraborty,T.,
Charbit,A., Chetouani,F., Couve,E., de Daruvar,A., Dehoux,P.,
Domann,E., Dominguez-Bernal,G., Duchaud,E., Durant,L.,
Dussurget,O., Entian,K.D., Fsihi,H., Portillo,F.G., Garrido,P.,
Gautier,L., Goebel,W., Gomez-Lopez,N., Hain,T., Hauf,J.,
Jackson,D., Jones,L.M., Kaerst,U., KrefT,J., Kuhn,M., Kunst,F.,
Kurapatk,G., Madueno,E., Maitournam,A., Vicente,J.M., Ng,E.,
Nedjari,H., Nordsiek,G., Novella,S., de Pablos,B., Perez-Diaz,J.C.,
Purcell,R., Rammel,B., Rose,M., Schlueter,T., Simoes,N.,
Tierrez,A., Vazquez-Boland,J.A., Voss,H., Wehland,J. and Cossart,P.
Comparative genomics of Listeria species
Science 294 (5543), 849-852 (2001)
21537379
PUBMED 11679669
REFERENCE
2 (bases 1 to 324050)
AUTHORS Glaser,P., Frangeul,L. and Rusniok,C.
DIRECT SUBMISSION
TITLE Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
JOURNAL Cedex 15, FRANCE
COMMENT E-mail: pglaser@pasteur.fr
FEATURES
source Location/Qualifiers
1. 324050
/organism="Listeria monocytogenes"
/strain="EGD-e"
/db_xref="taxon:1639"
complement(290..492)
/gene="lmo2321"
complement(290..478)
/gene="lmo2321"
/codon_start=1
/transl_table=11
/protein_id="CAD00399.1"
/db_xref="GI:16411810"
/db_xref="SPTREMBL:Q928W7"
/translation="MAEPIFKQTIFGNSEIFIDRTKMIANPAFRQKIALIETGCEK
MTDYIELKLKGYEEVTR"
complement(487..492)
/gene="lmo2321"
complement(586..801)
/gene="lmo2322"
complement(586..801)
/gene="lmo2322"
/codon_start=1
/transl_table=11
/product="gp44 [Bacteriophage A118]"
/protein_id="CAD00400.1"
/db_xref="GI:16411811"
/db_xref="SPTREMBL:Q8Y4V8"
/translation="MSEVLVSASYEGYKSKINSFTENNIVKRFKKIDVEVRKRAE
TFNKKYEVTRKELVDGHLRELIIPRTL"
complement(join(798..1331,1340..1345))
/gene="lmo2323"

complement(798..1331)
/gene="lmo2323"
/codon_start=1
/transl_table=11
/product="gp43 [Bacteriophage A118]"
/protein_id="CAD00401.1"
/db_xref="GI:16411812"
/db_xref="SPTREMBL:Q8Y4V7"
/translation="MFQKSTSAAMQVLAETRTOKELATDSYVTPALISNOIKGKRT
VLSQAEQLIDSYNEPSTYLFPAHEFSNGMIPPLFDGLDNNHASLNNRFELEVEEAMN
TLKNGLETFTNLRKGDMLQREAAKQAISETIDTVASALTNLNASTAIATFNIDLCQVLN
KRDLYYQKSGLVRSCGK"
complement(810..815)
/gene="lmo2322"
complement(810..815)
/gene="lmo2322"
complement(810..815)
/gene="lmo2322"
complement(1340..1345)
/gene="lmo2323"
complement(1454..2230)
/gene="lmo2324"
complement(1454..2230)
/gene="lmo2324"
/notes="similar to anti-repressor [Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00402.1"
/db_xref="GI:16411813"
/db_xref="SPTREMBL:Q8Y4V6"
/translation="MSNLOVIANEMLPVLNEOGEKFXVDARMLHEKLLVNTFRNDWIY
RMIGNTGYENGLDFYSTLKTNGRPSLTNYFLTDLTAKELAMVQNNEMGRAIRKFIETV
EKQARKLAYEPAFSYMIEDPVARAKKWIIEQOEKVLKLEEQPKVVFABAVQTS
ENTILVKLATILKQKGLDIGNRLFEWLRGSGYLLSKGAYYNKPSQKAMNIGLFEQK
THIHDRNGLMKTTYTPQVTKGQGVYLLNKLLEHNQVII"
complement(2212..2487)
/gene="lmo2325"
complement(2212..2487)
/gene="lmo2325"
/codon_start=1
/transl_table=11
/protein_id="CAD00403.1"
/db_xref="GI:16411814"
/db_xref="SPTREMBL:Q8Y4V5"
/translation="MIFTVNSFPQSGHVFTPGTLVTNTEHSGOLPSPVPELIISCFALF
RLCSLICLSNFFESNLILPILPTSLHKNYSTVKGRTERTKQIYK"
complement(2236..2241)
/gene="lmo2324"
complement(2236..2241)
/gene="lmo2324"
complement(join(2493..2774,2782..2787))
/gene="lmo2326"
complement(2493..2774)
/gene="lmo2326"
/notes="similar to protein gp41 [Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00404.1"
/db_xref="GI:16411815"
/db_xref="SPTREMBL:Q8Y4V4"
/translation="MNNIKQAIILKLETILENGNAIESGSPVKYSVTKNLSLLEKDOE
LKIEMKVELNGVEDSIENATLLEKRLIEAKSLVEDLASTINSLEIKVK"
complement(2498..2503)
/gene="lmo2325"
complement(2498..2503)
/gene="lmo2325"
complement(2782..2787)
/gene="lmo2326"
complement(2800..3097)
/gene="lmo2327"
complement(2800..3084)
/gene="lmo2327"
/codon_start=1
/transl_table=11
/protein_id="CAD00405.1"

```

```
QY 75 ArgGlyPheTyrLysAspAlaSerThrIleArgLys-----AsnSerArgIleSer 91
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 87104 AAAGGTTCTTTGAAGACGACAAC-----CGTAAAGCCGCATACGACAAAGATTAGC 87051

QY 92 ArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 87050 ACGTTAAAGAATATATTACGAACTCTGCACTTTGGCTGCTCTATTTTGTTATT--- 86994

QY 112 AsnProAsnProArgAspIleLys-----
:::|||||
Db 86993 -----AAAAAGTAAATAGTAAATGAGGTAATTTGGAATTCGAAGATTATATA 86946

QY 120 -----PheAspAspGluArgProHisLysArgAlaLysSerArgSerLysSerGlnSer 137
||||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 86945 AGCGTATTTAAATTGATTAGACGGCACAATGTATCGCGGACGAGAAGTTATCCCGAAGC 86886

QY 138 Ser 138
:::
Db 86885 AAT 86883

RESULT 17
AX417048/c 349980 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 4039 from Patent WO0228891.
ACCESSION AX417048
VERSION AX417048.1 GI:21449658
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
REFERENCE 1
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 4039 11-APR-2002;
Pasteur Institut (FR)
FEATURES
Location/Qualifiers
1..349980
/organism="Listeria innocua"
/db_xref="taxon:1642"
/note="seq 2058, original length: 3.011.208 replaced
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:
0.900.001 to 1.249.980-seq 4035: 1.200.001 to
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:
1.800.001 to 2.149.980-seq 4038: 2.100.001 to
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"
BASE COUNT 101702 a 72631 c 57863 g 117784 t
ORIGIN

Alignment Scores:
Pred. No.: 2e-11 Length: 349980
Score: 248.00 Matches: 63
Percent Similarity: 57.76% Conservative: 30
Best Local Similarity: 39.13% Mismatches: 35
Query Match: 24.58% Indels: 34
DB: 6 Gaps: 7

US-09-868-352-23 (1-192) x AX417048 (1-349980)
QY 2 ArgLysGluVal---ThrProGluMetLeu-----AsnTyrAsnLys 14
|||||
Db 124920 CGAAAGAAGTATATACACAGAAATTATCCGAGATATTTAAATGGAAACTCAAAAA 124861

QY 15 TyrProGlyProGlnPheIleHisPheGluAsnIleValLysSerAspIleGluPhe 34
|||:::
Db 124860 ATGTACGACAGGAGGA-AGACACGCTG-----ACGATTACGATTCAAGATTAAACTAC 124808

QY 35 GlnLeuValIleAsnGlnLysSerAlaPheAspValThrValPheGlyGlnArgPheSer 54
:::|||||
Db 124807 GAAATTATCACAATTATCGCGATGCTTTTGACGAGAAAATTGAATGAACGATTTAGC 124748

QY 55 GlulleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeu 74
```

```
Db 124747 GATATTCTTGGCGGATGATTATATAGTAGCGGCTTACGATCAACTTAGACTG 124688

QY 75 ArgGlyPheTyrLysAspAlaSerThrIleArgLys-----AsnSerArgIleSer 91
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 124687 AAAGGTTCTTTGAAGACGACAAC-----CGTAAAGCCGCATACGACAAAGATTAGC 124634

QY 92 ArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 124633 ACGTTAAAGAATATATTACGAACTCTGCACTTTGGCTGCTCTATTTTGTTATT--- 124577

QY 112 AsnProAsnProArgAspIleLys-----
:::|||||
Db 124576 -----AAAAAGTAAATAGTAAATGAGGTAATTTGGAATTCGAAGATTATATA 124529

QY 120 -----PheAspAspGluArgProHisLysArgAlaLysSerArgSerLysSerGlnSer 137
||||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 124528 AGCGTATTTAAATTGATTAGACGGCACAATGTATCGCGGACGAGAAGTTATCCCGAAGC 124469

QY 138 Ser 138
:::
Db 124468 AAT 124466

RESULT 18
AX142599
LOCUS
DEFINITION Sequence 1321 from Patent WO0134809.
ACCESSION AX142599
VERSION AX142599.1 GI:14282171
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 390)
AUTHORS Kimmerly, W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 1321 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
Location/Qualifiers
1..390
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic nucleic acid sequence"
BASE COUNT 158 a 41 c 67 g 124 t
ORIGIN

Alignment Scores:
Pred. No.: 9.32e-15 Length: 390
Score: 243.50 Matches: 45
Percent Similarity: 68.32% Conservative: 24
Best Local Similarity: 44.55% Mismatches: 29
Query Match: 24.13% Indels: 3
DB: 6 Gaps: 2

US-09-868-352-23 (1-192) x AX142599 (1-390)
QY 26 IleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAsp 45
:::|||||
Db 1 ATGATTAAAGTCGACCAACAATATTTTGAATTGATAGAGAATATAGAGAATCTTTTGTAT 60

QY 46 ValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGly 65
:::|||||
Db 61 GAGGAATATTTTTCAGCTAGGTATTCGGATATATTCGGAATATATGACAAATATGATTATCGTAGGT 120

QY 66 AspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArg 85
|||:::|||||
Db 121 GACTATGTTACGATCAATTTACCTTAAAGAGATTTTATAAAGATAGTATAATAAAGGCA 180

QY 86 Lys---AsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGly 104
:::|||||
Db 181 GAAATAAGTAACGATTTTCAAGTATACAAGATTATATATACTAGAATATTCGTAATTTTGGT 240

QY 105 CysAlaTyrPheValLeuGluAsnProAsnProArgAspIle-----LysPheAspAsp 122
```

```
/transl_table=11
/protein_id="CAC97562.1"
/db_xref="GI:16414858"
/db_xref="SPTREMBL:Q929E2"
/translacion="MNJHURFYVTLAHMEHYTKAAEKLLITOPSLSHATISSLEQELG
IPLEKGRNIGLSKAGKGVYVEESLMDIDAGVATTKKAANGEGQIDLAFTQTLGT
IAVPLKQVEITQTEPEKKIDVFHTGVSIDIIGLEKKKYDIGICSLDENRNHFPTP
SAKQELVLPKSHPLAEKOWIDLIEVPPHIAFSKKSGLRPIIDDLERKIGDYKI
AYETEVDVITAGVAQNFGLAVPNMFLINLVVDVKVLPISPNWEREFYLAIVKQNL
TPAERKQWLEHRI"
8543. .9397
/gene="lin2336"
CDS
8543. .9397
/gene="lin2336"
/notes="similar to unknown proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC97563.1"
/db_xref="GI:16414859"
/db_xref="SPTREMBL:Q929E1"
/translacion="MTNANGDLKKCPITISSYTLGTEVSPERVRIAAENGFDGIGLR
AENVYDALAGLTDENVRLDEHNIRKVEVEYITQMGTAEDRTVEOKKEOTTFHMA
RLFGVKNIGLLEKIPBEOIITALGELCDRAEELITGLEFMPYGVADIAAARVRAE
ACGRDQAQLCDTWHHAROKTAEISKNPADRIYSIQLCDVHETPYKELREESLHNR
LAQEGYGDVTFGARILKEHGVNPRVMGVEISDSMVETGLEVAAIKVYNATKKVLDE
AWPEISPK"
9467. .11392
/gene="lin2337"
CDS
9467. .11392
/gene="lin2337"
/notes="similar to NADH oxidase"
/codon_start=1
/transl_table=11
/protein_id="CAC97564.1"
/db_xref="GI:16414860"
/db_xref="SPTREMBL:Q929E0"
/translacion="MKFNSMFSPIDIGPMRVNPNRVFVSPMCMNYANTDGLTDTSLAY
YKRALGFGGLITEPTVDVVRAGKAGKACLYSDHOIASFKRVIDACHDAGAKISVO
LOHAGPENSXSGVPLKAAASATASAEGRNTPPAISREELYELIELYGEALPAKKAQ
ADAVEICANGLYSVSLSGRTNKRVDGCGCFENRMRPLRIIESIRKRVGHISAIL
CRVNTQDGLSVQSDATVAAYLEDGLDGLHVSRSVHRIEDYMWAPITVHLHAGFSS
DLNTOIKRAVYIPVITVGRTEPHYAEALVREGRAVLAFGROSLADPPTPNKAAAGK
IDELLPCACLOGVANNYAGKPTICLVNPLLGRESEAYLPKTPSKKVYVIGGVGL
YAGMAGSRGHDVTVYEASDIIGOMRLAAYPGKGLDITNNRSYIKCKQDFDVEIKT
NTPVTPELIDIAPDVAVIATGATPLVLPVLPGLIEDSGLIHAVLDLDCESKSGKVLVY
GGMVGSETAAFLGEAGHDYTVVLELREDEGVADYSERKFLMRDFDDEYKIKSTINAKV
TSFLPDGVTYSDEKEVIRIDGDFSVVLAMGSRAYNPLEEAIRKIVPVTIVIGDAVRA
RRALDATKEALDAVLQL"
Alignment Scores:
Pred. No.: 1.29e-11 Length: 248050
Score: 248.00 Matches: 63
Percent Similarity: 57.76% Conservative: 30
Best Local Similarity: 39.13% Mismatches: 35
Query Match: 24.58% Indels: 34
DB: 1 Gaps: 7
US-09-868-352-23 (1-192) x AL596172 (1-248050)
Qy 2 ArgLysGluVal---ThrProGluMetLeu-----AsnTyrAsnLys 14
|||||
Db 162920 CGAAAAGAAGTATATACACCAAGATTACTCCGAGATATTTTAAATGCAAACTCAAAAA 162861
Qy 15 TyrProGlyProGlnPheIleHisPheGluAsnIleValLysSerAspIleGluPhe 34
|||||
Db 162860 ATGTACGACAGAGAGGA-AGACACGTG-----ACGATTACGATTCAAGATTAACTAC 162808
Qy 35 GlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPheSer 54
|||||
Db 162807 GAAATTATACCAATATTCGCGATGCTTTTCACGAGGAAAAATTTGAATGAACGATTATAGC 162748
Qy 55 GluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluInLeuArgLeu 74
|||||
```

```
Db 162747 GATATTTCTGGCGCATATGATTATATAGTAGCGACTGGGCTTACGATCAACTTAGACTG 162688
Qy 75 ArgGlyPheTyrLysAspAlaSerThrIleArgLys-----AsnSerArgIleSer 91
|||||
Db 162687 AAAGGTTTCTTTGAAGACGACAAC-----CGTAAAGCCGATACGACAAACGATTAGC 162634
Qy 92 ArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111
|||||
Db 162633 ACGTTAAAGAATATATTTAGCACTACTGCACTTGGCTGCTTATTTGTTATTT--- 162577
Qy 112 AsnProAsnProArgAspIleLys----- 119
|||||
Db 162576 -----AAAAAGTAAATAGTAATAGGTAATTTGGAATTGAAGAAATTATA 162529
Qy 120 -----PheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSer 137
|||||
Db 162528 AGCGTATTTAATGATTGATTAGACGACACAATGATCGCGGACGAGAAGTTATCCCGAAGC 162469
Qy 138 Ser 138
|||||
Db 162468 AAT 162466
RESULT 16
AX413018/c 349980 bp DNA linear PAT 14-JUN-2002
LOCUS AX413018
DEFINITION Sequence 9 from Patent WO0228891.
ACCESSION AX413018
VERSION AX413018.1 GI:21445476
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
REFERENCE 1
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 9 11-APR-2002;
Pasteur Institut (FR)
FEATURES
source
1..349980
/organism="Listeria innocua"
/db_xref="taxon:1642"
/notes="old seq 9 (684707) splitted in 3-new seq 9 = old
seq 9 from 1 to 349980-seq 4027 = old seq 9 from 300000 to
649980-seq 4028 = old seq 9 from 600000 to 684707"
BASE COUNT 104581 a 69113 c 62414 g 113872 t
ORIGIN
Alignment Scores:
Pred. No.: 2e-11 Length: 349980
Score: 248.00 Matches: 63
Percent Similarity: 57.76% Conservative: 30
Best Local Similarity: 39.13% Mismatches: 35
Query Match: 24.58% Indels: 34
DB: 6 Gaps: 7
US-09-868-352-23 (1-192) x AX413018 (1-349980)
Qy 2 ArgLysGluVal---ThrProGluMetLeu-----AsnTyrAsnLys 14
|||||
Db 87337 CGAAAAGAAGTATATACACCAAGATTACTCCGAGATATTTTAAATGCAAACTCAAAAA 87278
Qy 15 TyrProGlyProGlnPheIleHisPheGluAsnIleValLysSerAspIleGluPhe 34
|||||
Db 87277 ATGTACGACAGAGAGGA-AGACACGTG-----ACGATTACGATTCAAGATTAACTAC 87225
Qy 35 GlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPheSer 54
|||||
Db 87224 GAAATTATACCAATATTCGCGATGCTTTTCACGAGGAAAAATTTGAATGAACGATTATAGC 87165
Qy 55 GluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluInLeuArgLeu 74
|||||
Db 87164 GATATTTCTGGCGCATATGATTATATAGTAGCGACTGGGCTTACGATCAACTTAGACTG 87105
```





```

SSAFILMFALFNVBEVLESIPIDFRILMIIGALVSSITAIIPMPFGKPFLSQVETTWI
3548.3892
/gene="SAV0624"
CDS
3548..3892
/gene="SAV0624"
/translation="similar to Na+/H+ antiporter"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56786.1"
/db_xref="GI:14246393"
/translation="MNLLILVIGLVFICGTMYMILSINLIRIVIGISIVTHAGNLIIM
SMGYGSSRSEPLITGQNQLFVDPDLQALVLAIVTAVICFGMTAFLLVLVRYTKYKED
EIEGLRGEDAK"
3882..5378
/gene="SAV0625"
CDS
3882..5378
/gene="SAV0625"
/translation="similar to Na+/H+ antiporter subunit"
/codon_start=1
/transl_table=11
/product="MnhD homologue"
/protein_id="BAB56787.1"
/db_xref="GI:14246394"
/translation="MLSNLLILPMLLPFLCALILVFLKNDRISKLYLYLGTMTITII
SLMLIYVQRHPITLDGGSAPFGIOFLGDSLSLIMVTASFVITLIMAYGFGCE
HKARYHLPSFELFSLVGVISFLTDLFNLVYMFEMLLASFLVILGQSVQOLRAA
IIVVLNIQSWFLGIGLLYKTVGNFNSHIAIRLNDMDNRTVMISLILFVAFS
AKALVFLMIPKAYAVLNTLAAFLAALMTKVAYALIRFFTLFDQHNDLHPLIA
TMAATTVIGAIQVIAIKYKIAAYQVVISIGFVILGLGTNTFAGINGAIFVLVNDI
VMTLFFIIGSVITGYQYVYVNLAKKEPLFGVAFIIMTFAIGVPPFSGFGPK
VLIFQGLGVNIXGLALMILITSLIAMYSIFRFFYMYFGDKGEEVNEFKKIPLYRR
ILSLVYVYVIAIGIAAPVLNVTSDELNTSDQLYOKLVNPHLKGD"
5379..5861
/gene="SAV0626"
CDS
5379..5861
/gene="SAV0626"
/translation="similar to Na+/H+ antiporter"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56788.1"
/db_xref="GI:14246395"
/translation="MNOIVLNIILFVLVLFQEDHFKFSTFGYLGILVILVLR
FSDDFYVRKIWAIRKELGYVLQIITSSISTYINILFKTKDMNPGLLSYETRLTSDM
ATPLTLIIITFESTVIRISQDSKFFIHSIDVSEKESLRSIKHIEDLILEVSR"
5858..6160
/gene="SAV0627"
CDS
5858..6160
/gene="SAV0627"
/translation="similar to Na+/H+ antiporter"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56789.1"
/db_xref="GI:14246396"
/translation="MIOTITHIMIISLIIFGIALIICLFRLLKPGTTPADRVVTFDTT
SAVMSVIGVLSVLMGTVSFLDSIMLIAIISFVSSVISIRFIGGHVFNNGNKRNL"
6135..6572
/gene="SAV0628"
CDS
6135..6572
/gene="SAV0628"
/translation="conserved hypothetical protein"
/protein_id="BAB56790.1"
/db_xref="GI:14246397"
/translation="MEITKEIFSLIAAVMLLGSFIALISAIGIVKFDQVFLRSHAT
KSSTLSVLLFLIGLVLIYFVNTGFFSVRLLSLVFINLTSFVGHHVLAARYNGAYM
YRNDATHASILLSSNEQNSTEALQRAKKRREHKKWYQND"
6912..8954

```

```

/gene="SAV0629"
6912..8954
/gene="SAV0629"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB56791.1"
/db_xref="GI:14246398"
/translation="MEIFETILFIADVILSSVFTPIPKVPLAFIOIFLGMLLFITP
IPVQNFDESEFMVTIAPLLFVEGVNVSRLHURKIKPVMMLGALGVITTVICVGLF
IHWIPDLPICAAFAIAAIICTPDVAVAQAITTKVLPKGAITILEGESLLNDAAGII
FTIAGVLTVGAFSLVDVQLFIAAGVGLIGMALVRFRLTMRGYENIM
SFIIQTLTPVTYLIARLFAELHAGSIIAAVAGLVHGFDRIMOVRTLOKSYNHTNI
LGVYNGFVSEILGFLVPEVIKIITKEPHNLIPLIGITIVVALAVYLFREVVYVLY
PYELAISFPQKMTKNDNDNPTTEKPKRSLYALIMTLCGVHCTISALATLTPYEL
AGHAFTYRNDLLFIASGMVILSLVAQVLLPLITKPAKTYIGNMSEFKVARTILEQ
VIDLNOKSTFETSKYGNVKEYHDKLTKVEKDDENSEKELERLOKIAFNVEYETK
LESIVDEGQITNSVLENYRYAERTQVYQASLIRRMVILLGALLRRVQYRVNSAS
SLSVTDNLMELNKLNVHNVYVSRLSKETDKNTLEIGMVCDCYLMRIENLTPSNFF
NSASEDTITKILNALREQRILRELIDTDEVSEGTALKLREAINYDEMVIIVDSMT"
complement(9062..10381)
/gene="tnp"
complement(9062..10381)
/gene="tnp"
/translation="transposase for IS1181"
/protein_id="BAB56792.1"
/db_xref="GI:14246399"
Alignment Scores:
Pred. No.: 1.43e-11 Length: 348527
Score: 249.50 Matches: 50
Percent Similarity: 62.28% Conservative: 21
Best Local Similarity: 43.86% Mismatches: 30
Query Match: 24.73% Indels: 13
DB: 1 Gaps: 2
US-09-868-352-23 (1-192) x AP003360 (1-348527)
Qy 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAla 43
Db 276130 AAGAAATTGATAAAAGTAGACAATCTACTTTGAATTAATAGAAAATATCGCGAATGT 276189
Qy 44 PheAspValThrValPheGlnArgPheSerGluIleLeuLeuLysTyrAspPheIle 63
Db 276190 TTTAATGAAGAACAAATTTATTGCTAGGTATTTCAGATATTTTAGACAAATATGATTACATT 276249
Qy 64 ValGlyAspTyrGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSer--- 82
Db 276250 GTTGGTGACTATGGTTACGATCAATACGATTAAGGCTTTTACAAGATTTCTAATAAA 276309
Qy 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102
Db 276310 AAACGACAGATGACATAACGGTTTTCAAATATTCAGATTTACATATTTGAATATTGTAAC 276369
Qy 103 PheGlyCysAlaTyrPheValLeu----- 110
Db 276370 TTTGGTTGCTCTTACTTGTATTAGACATTTGTCTAAACAGAGGTTTAAAGTTAATC 276429
Qy 111 GluAsnProArgAspIleLysPheAspAspGluArg 124
Db 276430 GAAGAAGTTTCATCGCTGTGATGTAGATGACGACACAATAA 276471
RESULT 15
AL596172/c
LOCUS
DEFINITION Listeria innocua Clip11262 complete genome, segment 10/12.
ACCESSION AL596172
VERSION AL596172.1 GI:16414852
KEYWORDS
SOURCE Listeria innocua.

```





```

/protein_id="BAB94399.1"
/db_xref="GI:21203700"
/translation="MIIYRQVHHEGAPVEIITKTFQHSVKICDDSFSDTEIFKLLSL
LQDDIDHMKVS"
complement(6293..7123)
/gene="MW0535"
complement(6293..7123)
/gene="MW0535"
/note="ORFID:MW0535
hypothetical protein, similar to phosphomethylpyrimidine
kinase, thid homologue"
/codon_start=1
/transl_table=11
/protein_id="BAB94400.1"
/db_xref="GI:21203701"
/translation="MALKKVLTIAGSDTSAGAGMQADLKTFOELDTYGMVALTAIVTM
DKDTWSDHVDPLPMDVFKQLETALSIGPDAIKTGMGTETIIKRAGEVYEASNAQYF
VVDPMVCKGEDEVLPNGTEAMIKYLLPKATVTPNLFEGQLSLGKLNSIDMKK
AATIFLGAQHVIIKKGKALDQDKSYDLIDGQTFYQLTIDMFQOSYNHAGCCTFAA
ATTIYLDANGSPKEAVISAKAFVASAIKNGWMKNDVFVGVDPDHGAYNRIEHDIVETVEY
"
7307..7963
/gene="ung"
7307..7963
/gene="ung"
/note="ORFID:MW0536"
/codon_start=1
/transl_table=11
/product="uracil-DNA glycosylase"
/protein_id="BAB94401.1"
/db_xref="GI:21203702"
/translation="MENSQIFHDITTKHDFKAMHDFLEKEYSTAIVYDPRENIYAQFD
LTPFNKIVILGQDPYHPGPNQAGHLAFSVQPNAPKPPSLRNMYKELADDIGCVRQTP
HQQWARGELVLTVLTVRQGEANSHRDIGHWETFDIIRKAVSDYKHHVPIILWKP
AQQIKLIDTSKHCIIKSVHPSPLSAYRGGFGSKPYKANAYLESVGRKSPINWCESEA
"
7964..8344
/gene="MW0537"
7964..8344
/gene="MW0537"
/note="ORFID:MW0537"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB94402.1"
/db_xref="GI:21203703"
/translation="MLNRETLIARIEQELVQAEQAQHDHDFEKHYAIHILTSYAST
SNTPHIGEQMMRRRIANHMOPSOITQPTHQVTAETAEMGCKVNTTHSAHHNKSYS
OPSNQORLATDDDIGNGESIFDF"
8477..8845
/gene="MW0538"
8477..8845
/gene="MW0538"
/note="ORFID:MW0538"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94403.1"
/db_xref="GI:21203704"
/translation="MKLFILGALNMAVGTGAFGAHGLOGKIKSDHLSYLVWEKATTY
OMYHGLALLIGVISTGTSINYNWAGLIFAGIIFSGSLYLTLVLTQIKVLGATPIG
GVLFILGIMLLIATFKFAG"
8966..10450
/gene="MW0539"
8966..10450
/gene="MW0539"
/note="ORFID:MW0539"

```

## Alignment Scores:

Pred. No.:	1.13e-11	Length:	290150
Score:	249.50	Matches:	50
Percent Similarity:	62.28%	Conservative:	21
Best Local Similarity:	43.86%	Mismatches:	30

```

Query Match: 24.73% Indels: 13
DB: 1 Gaps: 2
US-09-868-352-23 (1-192) x AP004824 (1-290150)
QY 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAla 43
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 279243 AAGAATTTGATAAAAGTAGATCAACATTACTTTGAATTAAGAAATTTATCCGGAATGT 279302
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyAspPheIle 63
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 279303 TTTAATGAAGAACAAATTTATTCCTAGTATTACAGATATTTAGATAAATATGATTACATA 279362
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrlLysAspAlaSer--- 82
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 279363 GTTGGTGACTAGTTACGATCAATTACGATTAAAGGTTTTTACAAAGATTCTTAATAAA 279422
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrlLysGluTyrlCysAsn 102
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 279423 AAAGCAGAGATGAGTAAAGCTTTTCAAAATATTCAAGATTACATATTTGAATATTGTAAC 279482
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 103 PheGlyCysAlaTyrlPheValLeu----- 110
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 279483 TTTGGTTGCTTACTTTGTTTAAACACATTTCTTAACAACAGGTTAAAGATTAAATC 279542
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 111 GluAsnProAsnProArgAspIleLysPheAspGluArg 124
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 279543 GAAGAAGTTTCATCCGTCGTGATGTAGATGACGACAAATAAA 279584
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
RESULT 13
LOCUS AP003131 299050 bp DNA linear BCT 02-JUL-2002
DEFINITION Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
ACCESSION AP003131 BA000018
VERSION AP003131.2 GI:14349173
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,
Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M.,
Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,
Mizutani-Ui, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C.,
Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J.,
Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,
Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
JOURNAL Lancet 357 (9264), 1225-1240 (2001)
MEDLINE 21311952
PUBMED 11418146
REFERENCE
AUTHORS Aoki, K., Oguchi, A., Hosoyama, A., Nagai, Y., Kuroda, M., Hiramatsu, K.
and Kikuchi, H.
Direct Submission
Submitted (30-JAN-2001) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center; 2Chome 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@ente.go.jp, URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-8423, Fax:81-3-3481-8424)
On Jun 12, 2001 this sequence version replaced gi:13700445.
FEATURES
Location/Qualifiers
1..299050
/organism="Staphylococcus aureus subsp. aureus N315"
/strain="N315"
/sub_species="aureus N315"
/db_xref="taxon:158879"
248..1324
/gene="ilvE"
248..1324
CDS

```



[illegible]

```

DKLNGVLGDOGLAAWKPLFASAHLDAGAYFTVIIMGALAVIYAKLMLADITIKMPES
VPAVAKFLAIIPTTAALVIGLVIYIGLKNDSVINLITYIAEPFOILSONIFS
VLIVTFVSVEFEGHGNVLPVLDGIMGPLGLANNOVFOVHSGLEDLIAKAV
DKAHANGDYNVNWGSDWAFAGSGGTITLVIAIILFSRKDYKIVRGICLAPG
ININEVPLEGLPVVINAIFBFAVAPLISVIAIYATATAHLVDVNVNAPVWPTPI
MNFAMATGDEWRAIVLTINLIITFVIWVFVIAANKLEETELD"
gene
3527. .4963
/genes="bgls"
/notes="L179659"
3527. .4963
/genes="bgls"
/notes="L179659"
3527. .4963
/genes="bgls"
/notes="L179659"
/notes="EVIDENCE BY HOMOLOG Y BIO06.12 ENERGY METABOLISM.
Sugars. HYDROLYSIS OF TERMINAL, NON-REDUCING
BETA-D-GLUCOSE RESIDUES WITH RELEASE OF BETA-D-GLUCOSE"
/codon_start=1
/transl_table=11
/product="beta-glucosidase A (EC 3.2.1.21)"
/protein_id="AAK04273.1"
/db_xref="GI:12723027"
/transl_table=11
/translation="MEHKKLKAFNDPLMGSAAYOVEGAPEDGKKASVWDFVRI
PGKFTFGNGDVAVDHYHRYKEDVALMKELGLKSYRESIAWTRILPDGRGEVNAQGLK
FYEDLIDELIANEIEPIVTIYHMDLQALDLYGWNESREITADFYNAARVFNAPKRG
KVYVYSLNFTSQWLSLATHPPGKDKMLFYQVNHIANLTNAAVINKFHELEMD
QGQPSFATPYAKDSDDLNLVLAENAEELYSHFWDVLYGYEPTIAMAAYLRENGL
APFEAGDAELKSAKPDPLGNGIYQYATNANPVDGCVGSGFNTSGKGTTKESGI
PGLHKKIQNPFIDRTNWDWEIDPEGLRIALRITTSRYRIPVMITENGLEGYDKVEDGK
IHDDYRKLYLESYKALKEAMTDGAKVGYHTWSYTDLLSLWNGYKRYGVFVYVDQE
TMGSKLRPKDSYVYQHLIETNAEEV"
gene
5027. .5479
/genes="dut"
/notes="L181168"
5027. .5479
/genes="dut"
/notes="L181168"
/notes="EVIDENCE BY HOMOLOG Y BIO08.04 Pyrimidine
ribonucleotide biosynthesis. DE NOVO SYNTHESIS OF
THYMIDYLATE. BELONGS TO THE DUTPASE FAMILY"
/codon_start=1
/transl_table=11
/product="deoxyuridine 5'-triphosphate nucleotidhydrolase"
/protein_id="AAK04274.1"
/db_xref="GI:12723028"
/transl_table=11
/translation="MKIRGEVVTYKKNAGINIPKRSTHESAGYDIEAAETVSFAPGE
IKLIPGLKAYMAGEVLYMYDRSNPRKGLVINSVGVIDKDYNNPDNEGHMFEMQ
MRNFTDEEVIEKGERVGVGFVFPFLVADGDENQEXEERTGGFGSTGA"
5732. .7075
/genes="ybiB"
/notes="L181858"
5732. .7075
/genes="ybiB"
/notes="L181858"
/notes="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL.
General. similar to several 5'-nucleotidases"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK04275.1"
/db_xref="GI:12723029"
/transl_table=11
/translation="MKNRLTLHLNDLHSLERFPVIERFEAEKSHSEVESLFDLGN
VDVHPLTATEQCFNVLMNRLDLTAATIGNNEGILGTHMELSHLYDEANFPVLGN
LATNFQSEPLIVTTSFLRGIGLIGTAPVLAIPQAGWELSDPFVYVDELPLKDCD
FTLLSHLGNVDQIAKDYDEIDLIGATHHFLFEGAQVNOVTLAAAGRYGYVGQ
IDITFNNOLVDAQIEATNTLPQKSDLVETNGWELRGHKLLEGTVVADLYKALP
NVLPDFRGSYHLAKIFADYGNTKLCLMNSGLVTDLPNHLTDLDDLPNHLRILVR
FTTGBELRVLEIMVDSFLATOKIQMGFRGKTFGSLIFHGIIEPRGDFYIKNSE
NQSLEKISDEESQIVLPDQYLFANWYFPLKLKCKSEILFFPYFLREIVAEHFNK"
7090. .7845
/genes="ybiC"
/notes="L183216"
7090. .7845
/genes="ybiC"
/notes="L183216"
/notes="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL.
General. 53% identical to Yutd protein of B.subtilis"
/codon_start=1
/transl_table=11

```

```

/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK04276.1"
/db_xref="GI:12723030"
/transl_table=11
/translation="MAKVIDESKLNKYPGEHVMVAVGEVVQGFTHFTHVHNYREAF
DAKLQORSDVLKDYDIIVGDMGFEQLRLKGFSTSRKMLADNKIDHLEDEYVNEYC
NKGAYFIRIRTKDEAFVSEKLGFTKELKQGFDPKPRKRNRRNRNWARDQKVTED
KRSNSEARKKDFKIREKSTDRPKPVTKDNKYSVSKSOERKTDNKKQNPADSDOKK
FTMKRNPRKKVEQTETTKQENQGFVLRNKKD"
8039. .9136
/genes="ybiD"
/notes="L184159"
8039. .9136
/genes="ybiD"
/notes="L184159"
/notes="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL.
General. conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK04277.1"
/db_xref="GI:12723031"
/transl_table=11
/translation="WTENITETRPSTYGLTRDLIEWAIENGKEKKERATQVNDWLYR
KRVQSEEMSNLSAAFTDKLNEHFIPLNEQVVVQBSADGTXYLFMLPQMMIETVL
MRQSYGLSVCTVTVQCNMGCTFCASGLKKEKRDVTAAGEIVSQIMLVQKRYFDERGLDE
RVSHVVMGIGEPEDFNYEHLNMFRLVINDDGLAIGARHITVTCGCFMPAKIKEFAHE
NLQINLAISLHAPNNELRTSLMRTNAPLEKLEAIDYVTTNRRVTEYIMLSGE
NDSPEIAQGLADLIKPRNKLVSYNLIPYNPAEHIKYERSTKDKNTAKFYDVLKNGIN
CVVRQEQGLDIDAACQGLRSKQIKKNKAKLA"
9382. .9993
/genes="ybiE"
/notes="L185517"
9382. .9993
/genes="ybiE"
/notes="L185517"
/notes="EVIDENCE BY HOMOLOG Y BIO06.01 ENERGY METABOLISM.
Aerobic. similar to different oxidoreductases"
/codon_start=1
/transl_table=11
/product="oxidoreductase"
/protein_id="AAK04278.1"
/db_xref="GI:12723032"
/transl_table=11
/translation="MDYVLRNKRHRKKNFDGKKIPTVDVKQIISASLAPSAHNTQS
WHFVIVSEKREKREALLSEVDPGNHQIKAGAVIVLFSPTDLVERSDRARGAGELD
DEQLLFERNRYPMQYEGFELHESNLSINIGLVTNLVLAIKNYGESNIIMGFNRT
QKWNELLEVKEKRYRPELIIPLGNSDEKGPYSRLPQSQIMEIR"
BASE COUNT 3369 a 1669 c 1992 g 3068 t
ORIGIN

Alignment Scores:
Pred. No.: 2,58e-21 Length: 10098
Score: 331.50 Matches: 75
Percent Similarity: 51.83% Conservative: 38
Best Local Similarity: 34.40% Mismatches: 64
Query Match: 32.85% Indels: 41
DB: 1 Gaps: 5

US-09-868-352-23 (1-192) x AE006255 (1-10098)
Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
Db 7090 ATGCTAAAGTAATAGACGAAGTAAGTTAAATTAATATAATATCTCGGCAACATGTT 7149
Qy 21 IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40
Db 7150 ATGCACATAGGGGAAGTTGTTCAAGTTGGACACAGCACTTTTCATATTTTCATAATAT 7209
Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluLeuLeuLysTyr 60
Db 7210 CGTGAGCGCTTTTGATGCGGAAAAATTAGAACACACGCTTTTCTGATGTTTAGAAGTAT 7269
Qy 61 AspPheIleValGlyAspTyrGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
Db 7270 GATTATATTCGACAGATCGGGCTTTTGACCAATTCGTCAGAGGTTTTTTC----- 7323
Qy 81 AlaSerThrIleArgLys-----AsnSerArgIleSerArgLeuGluAspTyr 96

```





21357209  
11463916  
2 (bases 1 to 10029)  
Tetzelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Otterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E., Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.  
Direct Submission  
Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

**TITLE**  
JOURNAL  
Medical Center Dr, Rockville, MD 20850, USA

**FEATURES**  
Location/Qualifiers  
1..10029  
/organism="Streptococcus pneumoniae TIGR4"  
/strain="TIGR4"  
/db\_xref="taxon:170187"  
216..344  
/gene="Sp0763"  
216..344  
/gene="Sp0763"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK74901.1"  
/db\_xref="GI:14972238"  
/translation="MTIARFSRATESWNLGMVEISPCCLDFIQKSVTQALFKLLIH"  
478..1413  
/gene="Sp0764"  
478..1413  
/gene="Sp0764"  
/note="identified by match to PFAM protein family HMM PF0180"  
/codon\_start=1  
/transl\_table=11  
/product="dihydroorotate dehydrogenase A"  
/protein\_id="AAK74902.1"  
/db\_xref="GI:14972239"  
/translation="MVSTKTKIAGGFEDNCLMNAAGVACMTIEELVEVKNSAAGTFTV  
KTATLDFRQGPPEYQDPVGSINSMLPNNGLDYLDYLLDQEKESNRTFFLSLV  
GSPSEHTTILRKQESDFRGLTELNLSCPNPKGQIAYDFTTDRILAEVAFYTK  
PLGILPPYFDLVHFDQAAALFNKYPLKFKVNCVNSIGNGLYIEDESWTIRPKNGFGSI  
GSEYIKTPALANVAFVQRNPQIIGTGGVLTGRDAFEHILCGASMVQVGTLLHKE  
GVSAFDRTINELKALIMVEKGYSELDPRGLKLYID"  
1447..2484  
/gene="Sp0765"  
1447..2484  
/gene="Sp0765"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK74903.1"  
/db\_xref="GI:14972240"  
/translation="MLAIEESOKLTLSNLPSLSLFTGDOGEFVWKSQMLKQIGYDS  
ADLNPAFYDMKEVNVKQVELELSVLPFFADEKIVILDYFMDITAKKRLTDDLSKF  
EEDLQNPSPPTTKLIIFAEGKLDKSRRLVKLLRKDAKAFSAVEKEGLRQYFKQWSQF  
QGLQPTNHSFNFKLISGQFQSTKRLNFLLQSKVANSVLEEEDIYNALPKTLQDNIF  
DLTQPLTKMQDAQDLVRLDTLQGEDEIKLIAVMLGQFTFTQVRIKLAESGQTESQI  
ASSLSYLGRLNPYQIKFALRDSRGLSLSFLKQAIISLLEITDQIKTKLYEKGFLFE  
KALLOIASQVN"  
2657..3262  
/gene="Sp0766"  
2657..3262  
/gene="Sp0766"  
/codon\_start=1  
/transl\_table=11  
/product="superoxide dismutase, manganese-dependent"  
/protein\_id="AAK74904.1"  
/db\_xref="GI:14972241"

**translation="MAILPELPAYDALEPYIDAETMHLHDKHQTQYVNNANAALA**  
**KHPEIGDEALLADVESIPADIRQALINNGGHLHAFWELMTPEKTAPEAAALAA**  
**IDATYFGSFEFQAAATAATRTFGSGWMLVNNKEGKLEVTSTANDQDTPISECKKPI**  
**GLDWEHAYVVKYRNVRPDIKAFFSVINNKKVDLEYAAAK"**  
3418..3930  
/gene="Sp0767"  
3418..3930  
/gene="Sp0767"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK74905.1"  
/db\_xref="GI:14972242"  
/translation="MRKEIAPELYNNKRPPEFHLHGDKVETEGIAFSLVENIKDAF  
DVTTFNQRFSEVLTKFDYIVGDMSEQLRLRGFYKDRTEEKLEKISRLODYLLVEYS  
YGCAYFVLENEAPKRASFDKMKRTEETPSRKGAQPTQTKRKSNAADKKNRRQRKDDK  
GORHFVIRQK"  
3951..5036  
/gene="Sp0768"  
3951..5036  
/gene="Sp0768"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK74906.1"  
/db\_xref="GI:14972243"  
/translation="MRPSIHSLVHQTMOEWLEQGEKKFRADQIWEWLYRKRVRQSEF  
MTNLSKOLIAKLNDFQVNPVKQRIVOESADGTVKYLFELPDGMLEITVIMRQHYGLS  
VCVPTQVCNIGCTFCASGLIKKORLNNGEIVAQIMLVQKRYFDERGQDERISHTVVM  
GIGEPFDNYNNLNFRTIINDKGMAGIHRTVSTGLAHKIFADFGVGVNLAVS  
LHAPNELRSSIMKINRAPPIEKLFAAIEVYIETTRRVTFEYIMLNEYNDGVEQALE  
LTELKNIKLSLYNPLIPYNPVSEHQYSRSKPECVLAIFYDLTKKGVNCVVRQEHGT  
DIDAACQGLRSNTMKRDRQKAAVNP"  
5039..5518  
/gene="Sp0769"  
5039..5518  
/gene="Sp0769"  
/note="similar to GP:7110140; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="VanZf-related protein"  
/protein\_id="AAK74907.1"  
/db\_xref="GI:14972244"  
/translation="MRKTYNHILVMGVIFYSTICIVCFCTPQEQSTVGVTGPIQIHLG  
RLVFLTPFNSLMKLGESDIGQLCHIFLQNLINLVFLPFLIFQLLYLFPNLRKTKKY  
LLFSFLVSLGTIECTQLILDFFDFNRFVEIDDLWTNLGYLAWLLYKRLHKNVKN"  
5520..7061  
/gene="Sp0770"  
5520..7061  
/gene="Sp0770"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, ATP-binding protein"  
/protein\_id="AAK74908.1"  
/db\_xref="GI:14972245"  
/translation="MSILEVKNLSHGFGORAIPEDEVSRLLKGEHIGLVGANGECKST  
FMSVTGKMLPDECKVSEKSVYTVAGIDQHSVLAERSQVRDLRTAFDFELFKAERARIN  
DLYMKMGADVDALMEVEYGEQDLRESDFRTLDKAIIDEVARALGVDMFGMDTDT  
SLSGQDPTVYLLAKLLEKPDILLDEPTNYLDAEHIDWLKRYLQNYENAFVLIASHDI  
PFLNDVTNIYHVHENQOLTRYSGDYVQFOEYVAMKSOLEAAEYERQOKEIADLKDFVA  
RNKARVATRNAMSQKKLDKMDIIEIQSEKPKPSDFRPAFTPGTFIQAKNLQIGY  
DRPTKPLNLTFFERNQKVAIIIGANGIGKTTLLKSLGIIISPIAGEVERGDYLELGTFE  
QEVGGNRRQTPLEAVNAPALNOAEVRAALARCGLTTTKHIESQIQVLSGGQAKVRF  
CLLMNRENNVLDDEPTNHLDDVADKDELKALKEYSRLIMVCHPEPDFVEGDIQIDW  
FNFLT"  
complement(7117..7920)  
/gene="Sp0771"  
complement(7117..7920)  
/gene="Sp0771"  
/codon\_start=1

Db 106482 CATTACACGGGACAAAGTCGAAACGGAAGGATAGCTTTTCCCTGGTTGAAATAATC 106541  
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluLeuLeuLysTyr 60  
Db 106542 AAGATGCGCTTGTATGTGACGACTTTAATCAGCGTTTTTCAGAAGTATTACCAAGTTT 106601  
QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80  
Db 106602 GATTATATCGTGGGGACTGGAGCAACGACAGCTTCGCTACGAGGTTTTTACAAAGGAT 106661  
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100  
Db 106662 GACCGAACAGAGAAACCTGAAATAATCAGTCGTTTACAGACTACCTTTTAGAGTAT 106721  
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120  
Db 106722 TGTAGTATGTTGTGCGCTATTGCTTAGAAATGAAGCCCTTAAGCGAGCATCATTT 106781  
QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerLysSer 140  
Db 106782 GAC-----AAGAAATGCGTAGAGCGGAGAGAAACACCTTCTAGAAAAGGAGAA 106835  
QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160  
Db 106836 CCAGCTCAAACTAAACGA---AAGTCGATGCAGAT----- 106868  
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180  
Db 106869 AAGAAAAATAGACGCTGCTCAGAAAGACGACCATCTCAGAAAGAGGACAAAGAA 106922  
QY 181 SerAlaLysGlnHis-LeuLeuPheValArg 190  
Db 106923 -----CAAGCTCATTTGTCATTCGTCAGA 106947  
RESULT 7  
BD003783/C 5558 bp DNA linear PAT 31-JAN-2002  
LOCUS Polynucleotide of Streptococcus pneumoniae and sequence.  
DEFINITION BD003783  
ACCESSION BD003783.1 GI:18631744  
VERSION JP 2001501833-A/103.  
KEYWORDS unclassified.  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 5558)  
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C., Fannon,M. and Dougherty,B.A.  
TITLE Polynucleotide of Streptococcus pneumoniae and sequence  
JOURNAL Patent: JP 2001501833-A 103 13-FEB-2001;  
COMMENT OS Unidentified  
PN JP 2001501833-A/103  
PD 13-FEB-2001  
PE 30-OCT-1997 JP 1998520718  
PR 31-OCT-1996 US 60/029960  
PI CHARLES A KUNSCH,GIL H CHOI,PATRICK J DILLON,CRAIG A ROSEN, PI STEVEN C BARASH,  
PI MICHAEL FANNON,BRIAN A DOUGHERTY  
PC C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,  
PC C12N1/21,  
PC C12N5/10,C12P21/02,C12Q1/69,G06F17/30,C12N15/00,C12N5/00, PC G06F15/40  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source I..5558 /organism='Unidentified'.  
FEATURES Location/Qualifiers  
1..5558  
/organism='unidentified'  
/db\_xref='taxon:32644'  
BASE COUNT 1628 a 1239 c 1067 g 1624 t

ORIGIN  
Alignment Scores:  
Pred. No.: 9,29e-32 Length: 5558  
Score: 438.00 Matches: 88  
Percent Similarity: 68.89% Conservative: 36  
Best Local Similarity: 48.89% Mismatches: 50  
Query Match: 43.41% Indels: 6  
DB: Gaps: 4  
US-09-868-352-23 (1-192) x BD003783 (1-5558)  
QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20  
Db 4885 ATCGGAAAAGAAATTCACACTATAACAAGTTCTCTGCTCGGAGTTC 4826  
QY 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40  
Db 4825 CATTACACGGGACAAAGTCGAAACGGAAGGATAGCTTTTCCCTTGGTTGAAATAATC 4766  
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluLeuLeuLysTyr 60  
Db 4765 AAGGATGCGCTTGTATGTGACGACTTTTAAATCAGCGTTTTTCAGAAGTATTACCAAGTTT 4706  
QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80  
Db 4705 GATTATATCGTGGGGACTGGAGCAACGACAGCTTCGCTACGAGGTTTTTACAAAGGAT 4646  
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100  
Db 4645 GACCGAACAGAGAAACCTGAAATAATCAGTCGTTTACAGACTACCTTTTAGAGTAT 4586  
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120  
Db 4585 TGTAGTATGTTGTGCTATTGTTCTTAGAAATGAAGCCCTTAAGCGAGCATCATTT 4526  
QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerLysSer 140  
Db 4525 GAC-----AAGAAATGCGTAGAGCGGAGAGAAACACCTTCTAGAAAAGGAGAA 4472  
QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160  
Db 4471 CCAGCTCAAACTAAACGA---AAGTCGATGCAGATAAGAAA---AATAGACGCTGTCAG 4418  
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlu-----GlnAspLysGlu 178  
Db 4417 AAAGACGACAAAGGACAAACGCTCATTTGTCATTTCGTCAGAAAGTAGAGTAAGGAG 4358  
RESULT 8  
AE007384 10029 bp DNA linear BCT 31-AUG-2001  
LOCUS Streptococcus pneumoniae TIGR4 section 67 of 194 of the complete genome.  
DEFINITION AE007384 AE005672  
ACCESSION AE007384.1 GI:14972237  
VERSION AE007384.1  
KEYWORDS Streptococcus pneumoniae TIGR4.  
SOURCE Streptococcus pneumoniae TIGR4  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
REFERENCE 1 (bases 1 to 10029)  
AUTHORS Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Whittam,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umavam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D., Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T., Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.  
TITLE Complete genome sequence of a virulent isolate of Streptococcus pneumoniae  
JOURNAL Science 293 (5529), 498-506 (2001)

```
QVEGGNROTPLEAVNNAFFALNOAEVRAALARCGLTKKHIESQIOVLSSGEQAKVRF
CLLMRENNVLLVDEPTNHLVDVDAKDELKRALKEYRGSILMVCHEPDEFYECWIDQIWD
FNNLT"
gene      complement(7182. .7985)
          /gene="ppia"
CDS       /note="spr0679"
          complement(7182. .7985)
          /gene="ppia"
          /note="putative"
          /codon_start=1
          /transl_table=11
          /label=spr0679
          /product="Peptidyl-prolyl cis-trans isomerase"
          /protein_id="AAK99483.1"
          /db_xref="GI:15458267"
          /translation="MKLATLIIIIITVALACSSVQSRSLRGDDYVDSLSLAERSSKVA
AQAKEULNALTENANFQPSKEVDEAEVLIHTSQGDRIKLFKPLAPLAVENFL
THAKEGYINGITTHRVIDGFMVQGTGQGSIMWHDKTKDKGTGFKNEITPY
LYNIRGALAMANTQPNNGSQFFINQNSDTSKLPTSKYPOKIIIEAYKEGNPSLD
GKHPVFGQVIGGMDVVDKIAKAERKDEKDKPTTATTIDSEVVKDYDFKS"
gene      complement(8069. .8290)
          /gene="spr0680"
          complement(8069. .8290)
          /gene="spr0680"
          /codon_start=1
          /transl_table=11
          /label=spr0680
          /product="Hypothetical protein"
          /protein_id="AAK99484.1"
          /db_xref="GI:15458268"
          /translation="MSFYGLFYNGIAITPNTYLSAWFNVFIAALPLNLIVEPIARTI
LSSFKPFTGEEVEEVEDFDDDEIPTII"
gene      complement(8178. .8408)
          /gene="spr0681"
          complement(8178. .8408)
          /gene="spr0681"
          /codon_start=1
```

```
Alignment Scores:
Pred. No.:      8.27e-33      Length:      10029
Score:          452.50      Matches:      94
Percent Similarity: 66.49%      Conservative: 33
Best Local Similarity: 49.21%      Mismatches: 48
Query Match:     44.85%      Indels:      16
DB:              1          Gaps:      4
```

US-09-868-352-23 (1-192) x AE008445 (1-10029)

```
Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||
Db 3477 ATCGAAAAAGAAATTCACCTGAATTATACAACTATACAAAGTTTCCTGGTCCGGAGTTC 3536
|||||

Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
|||||
Db 3537 CATTTACACGGGACAGAGTCGAAACGGAAGGATAGTATTTTCTGTTGTAATAATC 3596
|||||

Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
|||||
Db 3597 AAGGATGCTTTGATGTGACGACTTTTAATCAGCGTTTTTCAGAAAGTATTAACCAAGTTT 3656
|||||

Qy 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
|||||
Db 3657 GATTATATCGTGGGGACTGGAGCAACGACAGCTTCGCCCTACGAGGTTTTTACAAGGAT 3716
|||||

Qy 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
|||||
Db 3717 GACCGAACAGAGAAAAAATCGTAAATAATCAGTCGTTTACAAGACTACCTTTTACAGTAT 3776
|||||

Qy 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
|||||
Db 3777 TGTAGTTATGGTGTGCGCTATTTTGTCTTAGAAAATGAAGCCCTTAAGCGGAGCATCATTT 3836
|||||

Qy 121 AspAspGluArgProHisLysArgLysSerArgSerLysSerGlnSerSerLysSer 140
```

```
|||||
Db 3837 GAC-----AAGAAATTCGTAAGACGGAAGAAACACCTTCTAGAAAAGGAAGAA 3890
|||||

Qy 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
|||||
Db 3891 CCGACTCAAACTAAACGA---AAGTCGAATGCAGAT----- 3923
|||||

Qy 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180
|||||
Db 3924 AAGAAAAATAGACGTCGTCAGAAAGACCAAGCATTCCTCAGAAAGAGACAAAGAA----- 3977
|||||

Qy 181 SerAlaLysGlnHis-LeuLeuPheValArg 190
|||||
Db 3978 -----CAACGTCATTTTGTTCATCGTCAGA 4002
|||||

RESULT 6
SPNEU1904      147006 bp      DNA      linear      HTG 11-JUL-2001
LOCUS      Streptococcus pneumoniae clone 654, *** SEQUENCING IN PROGRESS ***,
DEFINITION      in ordered pieces.
ACCESSION      AL449926
VERSION      AL449926.1. GI:11545151
KEYWORDS      HTG; HTGS_PHASE2.
SOURCE      Streptococcus pneumoniae.
ORGANISM      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
      Streptococcus.
REFERENCE      1 (bases 1 to 147006)
AUTHORS      Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
      Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
      Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
      and Garcia-Bustos,J.F.
TITLE      Annotated draft genomic sequence from a Streptococcus pneumoniae
JOURNAL      type 19F clinical isolate
MEDLINE      Microb. Drug Resist. 7 (2), 99-125 (2001)
PUBMED      21335329
REFERENCE      11442348
AUTHORS      2 (bases 1 to 147006)
      Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
      Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
      Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
      Garcia-Bustos,J.F.
TITLE      Direct Submission
JOURNAL      Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
      Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT      * NOTE: This is a 'working draft' sequence.
      * This sequence will be replaced
      * by the finished sequence as soon as it is available and
      * the accession number will be preserved.
FEATURES      Location/Qualifiers
      source      1..147006
      /organism="Streptococcus pneumoniae"
      /serotype="19F"
      /db_xref="taxon:1313"
      /clone="G54"
BASE COUNT      44762 a 26837 c 32084 g 43281 t 42 others
ORIGIN
Alignment Scores:
Pred. No.:      2.52e-31      Length:      147006
Score:          452.50      Matches:      94
Percent Similarity: 66.49%      Conservative: 33
Best Local Similarity: 49.21%      Mismatches: 48
Query Match:     44.85%      Indels:      16
DB:              2          Gaps:      4
US-09-868-352-23 (1-192) x SPNEU1904 (1-147006)
Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||
Db 106422 ATCGAAAAAGAAATTCACCTGAATATACAACTATACAAAGTTTCTCGGTCGGAGTTC 106481
|||||

Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
```

Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,  
Zook, C., Baltz, R.H., Jaskunas, S.R., Rostock, P.R. Jr., Skatrud, P.L.  
and Glass, J.I.

Genome of the bacterium *Streptococcus pneumoniae* strain R6  
J. Bacteriol. 183 (19), 5709-5717 (2001)

21429245

11544234

2 (bases 1 to 10029)

Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,  
DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,  
Gilmour, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., LaGace, R.,  
LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,  
McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,  
Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,  
Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,  
Zook, C., Baltz, R.H., Jaskunas, S.R., Rostock, P.R. Jr., Skatrud, P.L.  
and Glass, J.I.

Direct Submission

Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and  
Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA

Location/Qualifiers

1. .10029

/organism="Streptococcus pneumoniae R6"

/strain="R6"

/db\_xref="taxon:171101"

79. .185

/rpt\_family="RUP element"

/rpt\_type="dispersed"

/rpt\_unit="Rup\_A"

536. .1471

/gene="pyrDA"

/note="spr0672"

536. .1471

/gene="pyrDA"

/EC\_number="1.3.3.1"

/codon\_start=1

/transl\_table=11

/label="spr0672"

/product="Dihydroorotate dehydrogenase"

/protein\_id="AAK99476.1"

/db\_xref="GI:15458260"

/translation="MVSTKTIQAGFEFDCNLMNAGVACMTIEELKNSAAGTFVT

KTATLDFRGNPEPYQDPVPLGINSMLPNNGLDYLDYLDLDOEKESNRTFFLSLV  
GNSPETHILKKVQESDRGLTELNSCPNVPKQPIAYDFETDRILAEVAFYTK  
PLGIKLPFDIVHEDQAAINFKNIPKAFVNCVSSNGLYIEDSVVIRPNKGFGGI  
GGEYKPTALANWHAIFYQRLNQIQLIGTGGVLTGRDAFEHILCGASMVQVGTTLHKE  
GVSAFDRIITNELKAIMVERGESLEDFRGKLYID"

1505. .2542

/gene="hola"

/note="spr0673"

1505. .2542

/gene="hola"

/note="Putative"

/codon\_start=1

/transl\_table=11

/label="spr0673"

/product="DNA polymerase III, delta subunit"

/protein\_id="AAK99477.1"

/db\_xref="GI:15458261"

/translation="MLAIEESQKLTLNLSPLSLFTGTQGOGEVNMKSQVLKQIGYDS  
ADLNFAFDKVEVYKDVLELVLPFFADEKIVILDYDFMDITAKKRLTDDLKSF  
EYLDNNSPTTKLIIPAEGLDSKRRLVKLLKRDADVDAVEVEGELRQYFQKWSQK  
QGLOFNHSEFNLLKSGFQSETQKNLLFQSKANSVIEEDIVNAIPKTLQDNIF  
DITQFTLLTKMQARDLRLTLQGEDEIKLIAVMIGQFRTFTQVKILAESQTESQI  
ASSLSYLGRNPNPYQIKFALRDRSGLSLFKQALSYLIETDYQIKTLGYEKGLFE  
KALLQIASOVN"

2715. .3320

/gene="soda"

/note="spr0674"

2715. .3320

/gene="soda"

/EC\_number="1.15.1.1"

/codon\_start=1

/transl\_table=11  
/label="spr0674"  
/product="Manganese co-factored superoxide dismutase"  
/protein\_id="AAK99478.1"  
/db\_xref="GI:15458262"  
/translation="MAIILPELPYAYDALEPYIDAETMHLHDKKHQTQVNNANAALAE  
KHPEIGEDLEALADVESIPADIRQALINNGGHLNHALFWELMTPERTAPSAELAAE  
IDATFGSEFEQAATAATFRGSGWALVGVNKEGKLEVTSTANODTTISEGKKPIL  
GLDVMHAYVYKRYNRPDYIKAFESVINNRKVDLEYAAAK"

gene

3477. .4007

/gene="spr0675"

3477. .4007

/gene="spr0675"

/codon\_start=1

/transl\_table=11

/label="spr0675"

/product="Hypothetical protein"

/protein\_id="AAK99479.1"

/db\_xref="GI:15458263"

/translation="MRKEIAPELYNYNKFPPEFHLHGDKVETEGIAFSLVENIKDAF  
DYTFNQRESEVLTAKFDYIVGWSNEQLRGLFYKDDRTTEKLEKISRLODYLLEYCS  
YGCAYFVLENEAPKRASFQKMRKTEETPSRKGKPTQTKRKSNAADKKNRQKQDH  
SOKEDKQORHFVIRQK"

gene

4028. .5113

/gene="spr0676"

4028. .5113

/gene="spr0676"

/codon\_start=1

/transl\_table=11

/label="spr0676"

/product="Conserved hypothetical protein"

/protein\_id="AAK99480.1"

/db\_xref="GI:15458264"

/translation="MKPSIHSLAHOTMQEWLEQGEKKFRADQIEMWLYKRYQVSFEE  
MTNLSKDLIAKNDQFVNPLKQIRVDASDGTVKYLFELPDGMLTETVLMROHYGLS  
VGVTVQDCNIGCTFCASDLIKQRDLNNGELVAQIMLVQKYFAERQDERVSHIVVM  
GIGEPFDYNNVLFVCTINDDKGHAIGARHITVSTSLAHRINRFADEGVQVNLAYS  
LHAPNELRSSIMKINRAPPIEKLFPAIETIETNNRVTFEYIMLNEVNDVSVEQALE  
LAELKNIKLSYVNLIPYVSEHDOYSRSPKERVLAFTYDLTKRKGNCVVRQEYGT  
DIDAACGQLRSNTMRDROKAVAENP"

gene

5115. .5582

/gene="spr0677"

5115. .5582

/gene="spr0677"

/codon\_start=1

/transl\_table=11

/label="spr0677"

/product="Hypothetical protein"

/protein\_id="AAK99481.1"

/db\_xref="GI:15458265"

/translation="MTKKRELILRLGVAVYSLCIVYFCTPQBSQSTPGIQHLGRLVF  
LTPFNSLMKLGESVDIGQLCIFQNLINLNFLLIFQLLYLFPNLKRKTKVLIFS  
FLVSLGIECTQLILDFEFDENRVFEIDDLWTNTLGGYLAWLLYKRLHKNVKN"

gene

5584. .7125

/gene="ABC-NBD"

/note="spr0678"

5584. .7125

/gene="ABC-NBD"

/note="Putative"

/codon\_start=1

/transl\_table=11

/label="spr0678"

/product="ABC transporter ATP-binding protein - unknown

substrate"

/protein\_id="AAK99482.1"

/db\_xref="GI:15458266"

/translation="MSILEVKNLSHCGFGRDRAIFEDVSVFRLKGEHLGLVCANGEGKST  
FMSIVTGKMLPDEGVKSVYTAGVLDQHSVLAQRQSRVDRVLRDTAFDELFAEARN  
DLYMNAEDGADVALMEEVGELQDRLESDFYTLDAKIDEVARALGVMDFGMDTQVT  
SLSGGQRTVLLAKLLLEKPDILLDEPTNYLDAEHIDWLKRYLQYENAFVLISHDI  
PFLNDVINIVHVENQOLTRYSGDYQFOEVYAMKKSQLEAAAYEROQKEIADLKQFVA  
RNKARVATRNAMSROKKLDKMDIIELOSEKPKPSDFKPTPGREIFQAKNLOIGY  
NRLTPLNLTFERNOKVAILGANGIKGTTLLKSLGLIISPIAGEVERGDYLELGIFE

```

CDS      /gene="SpyM3_1152"
          complement(5346..5816)
          /gene="SpyM3_1152"
          /note="best blast hit: gb|AAK34295.1| (AE006583) putative
          repressor protein [Streptococcus pyogenes M1 GAS], and
          gb|AAL98085.1| (AE010066) putative repressor protein
          [Streptococcus pyogenes MGAS8232]"
          /codon_start=1
          /transl_table=11
          /product="putative repressor protein"
          /protein_id="AAM79759.1"
          /db_xref="GI:21904891"
          /translation="MKKSERLELIKMKVLTHPIETQHDLLRLLAELHGLTQATISRD
          PNETGIVKIPSGRIYIGLSQDSKRIKIVOGPRSIKSTILAVSDKTKGLEOHLVLYKV
          VNGSKLTKRYLLADFSKAIFSLIADDDSLLLIAKSPSEADMIRQEIILLWQGIT"
          complement(5803..6630)
          /gene="hlyA"
          /note="SpyM3_1153"
          complement(5803..6630)
          /gene="hlyA"
          /note="best blast hit: gb|AAL98086.1| (AE010066) putative
          hemolysin [Streptococcus pyogenes MGAS8232]"
          /codon_start=1
          /transl_table=11
          /product="putative hemolysin"
          /protein_id="AAM79760.1"
          /db_xref="GI:21904892"
          /translation="MPKERVDLVYKQGLPETREQAKRGVMAGLVVSVINGQRYDKPG
          DKINDGTELKLGKGLKYSRGGLKLEGLHFGVSVAOQIGIDIGASTGFTDVMLO
          DGAKLVADVGTNOLVWKURQDPRVRSMEOYNFRYAQVDFSEGVFASIDVSFLS
          LSLPLALHNVLSDQGVIALIKPQFAGREQIGKGIKVKDKQIHERKVIQKVMDFASG
          YGVTFKIDFSPIOGGHGNIETLAHLVKSQTPETLAPHLIQEVVAKHFEKHEKR"
          complement(6623..7495)
          /gene="fps"
          /note="SpyM3_1154"
          complement(6623..7495)
          /gene="fps"
          /note="farnesyl diphosphate synthase; best blast hit:
          gb|AAL98087.1| (AE010066) putative geranyltransferase
          [Streptococcus pyogenes MGAS8232]"
          /codon_start=1
          /transl_table=11
          /product="putative geranyltransferase"
          /protein_id="AAM79761.1"
          /db_xref="GI:21904893"
          /translation="MDKLARIDEAIRRYKTTNSGVSEELIDALLYSDSGKRIRPL
          ILLEMIEGVSLQNAHFDAALAEMLHTGSLIHDDLPAAMDNDYRGRLTNNHKQGE
          ATAILAGDSFLDPFGLIQAELNSEYKVALIQELSASGTFFGMVGMDMLDKGENQA
          LSLPQLSLIHLNKTGKLTTPFKAALVTDQAITVROOLEQAGMLIGHAFQIRDDILD
          VTASFDELGTPKPKDLFAERATYPSLLGLEASYQLLTESLDQALTIFOTLENDVGFKP
          QITPKLTGRLNA"
          complement(7495..7710)
          /gene="xseB"
          /note="SpyM3_1155"
          complement(7495..7710)
          /gene="xseB"
          /note="best blast hit: gb|AAK34298.1| (AE006583)
          exodeoxyribonuclease VII (small subunit), [Streptococcus
          pyogenes M1 GAS], and gb|AAL98088.1| (AE010066)
          exodeoxyribonuclease VII (small subunit) [Streptococcus
          pyogenes MGAS8232]"
          /codon_start=1
          /transl_table=11
          /product="putative exodeoxyribonuclease VII small subunit"
          /protein_id="AAM79762.1"
          /db_xref="GI:21904894"
          /translation="MSKTRTFEENLQDLETIVKNKLENGDVPLEEAFSEFQKMLLSKE
          LQKTLQAAENTLVKVMQADGTEVDD"
          complement(7688..9028)
          /gene="xseA"
          /note="SpyM3_1156"
          complement(7688..9028)

```

## Alignment Scores:

```

Pred. No.:      6.91e-42      Length:      50463
Score:           557.50      Matches:      107
Percent Similarity: 71.98%      Conservative: 42
Best Local Similarity: 51.69%      Mismatches: 41
Query Match:      1          Indels:      17
DB:              1          Gaps:      4

```

US-09-868-352-23 (1-192) x AE014159 (1-50463)

```

Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39611 ATGAAAAAGAAATTCGCCAGAAATGTACAACTATAATAATTTCCAGTCCCAATTT 39552
Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39551 ATTCAATTTGAAGCAAGTTAAGGCTGAAGCATTTGTTACTCTTAGAAGATGTT 39492
Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyr 60
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39491 AAGAACGCTTTTGATACGACTAGCTTTGTCACACGTTATACAGAAAGTCTCCTAAGATAT 39432
Qy 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39431 GACTATATTGTTGGCGATTGGGAAATCAACAGCTTCGCTCTTAAAGGCTTTTATAAGGAT 39372
Qy 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39371 AGTGATGATATCAAGAAACCAATCGCATCTCACGTTTAGAAGATTATATTAAAGATTT 39312
Qy 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39311 TGCATTTTGGTGTGCTTATTTCCTAGAAAATCTTCATCCACAGATATTAAATTT 39252
Qy 121 AspAspGluArgProHisLysArgArgLysSer----ArgSerLysSerGln----- 136
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39251 GAAGAGGAGCGCCACCAACCAAGACGAGAAAGAAATCACCTAATCAAAATCAATCGTCCGAG 39192
Qy 137 -----SerSerLysSerGlnThrArgAsnAsn 145
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39191 CCAACTATTCAATCAGCAGCCGCCAACACCTAGAGATTAATCGAAG---CGGTATCA 39135
Qy 146 ArgSerGlnThrAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39134 AAAGAAAGCAACCTGAAACCAACCAACCGTTTACCAGTCAAAACGTCGAAGTAATACTAAG 39075
Qy 166 ArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39074 CATAAAGAAAAGTCA---AAACGTAATCAGACTAGTCAACTTAATACCAAAATTTAGTCAT 39018
Qy 186 LeuLeuPheValArgLysAsn 192
    |||||.....|.....|.....|.....|.....|.....|.....|.....|
Db 39017 TTTATCATCAGAAAGAAAGAT 38997

RESULT 5
AE008445      10029 bp      DNA      linear      BCT 13-SEP-2001
LOCUS      Streptococcus pneumoniae R6 section 61 of 184 of the complete
DEFINITION      genome.
ACCESSION      AE008445
VERSION      AE008445.1 GI:15458259
KEYWORDS
SOURCE      Streptococcus pneumoniae R6.
ORGANISM      Streptococcus pneumoniae R6.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE      1 (bases 1 to 10029)
AUTHORS      Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
      Dehoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
      Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., Lagace,R.,
      Leblanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
      McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
      Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,

```

```

Qy 137 -----SerSerLysSerGlnThrArgAsn 145
Db 7071 CCAAACTATTCAAATCAGCAGCCGCAACACTAAGAGTAATCGAAG---CGGTATCA 7015

Qy 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
Db 7014 AAGAAAAGCAACTGAAAACGCGTTTACCAGTCAAAAACGTCGAAGTAATACTAAG 6955

Qy 166 ArgGlnGluArgHisLeLysGluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Db 6954 CATAAAGAAAAGTCA---AAACGTAAATCAGACTAGTCAACTTAATACCAAAATAGTCAT 6898

Qy 186 LeuLeuPheValArgLysAsn 192
Db 6897 TTTATCATCAGAAAGAAAGAT 6877

RESULT 4
AE014159/c
LOCUS
DEFINITION
  AE014159 50463 bp DNA linear BCT 19-JUL-2002
  Streptococcus pyogenes MGAS315, section 24 of 37 of the complete
  genome.
ACCESSION
  AE014159 AE014074
VERSION
  AE014159.1 GI:21904885
KEYWORDS
SOURCE
  ORGANISM
    Streptococcus pyogenes MGAS315.
    Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
    Streptococcus.
REFERENCE
  1 (bases 1 to 50463)
    Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
    Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
    Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
    Musser,J.M.
    Genome sequence of a serotype M3 strain of group A Streptococcus:
    phage-encoded toxins, the high-virulence phenotype, and clone
    emergence
    Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
  2 (bases 1 to 50463)
    Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
    Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
    Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
    Musser,J.M.
    Direct Submission
    Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
    Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
    Hamilton, MT 59840, USA
FEATURES
  source
    1. 50463
      /organism="Streptococcus pyogenes MGAS315"
      /strain="MGAS315"
      /serotype="M3"
      /db_xref="taxon:198466"
      /notes="group: A"
      complement(80..667)
      /gene="SpyM3_1147"
      complement(80..667)
      /gene="SpyM3_1147"
      /notes="best non-GAS blastp hit: gb|AAK04599.1|AE006285.7
      (AE006285) hypothetical protein [Lactococcus lactis subsp.
      lactis], and gb|AAK08221.1|AF320916_2 (AF320916) YfaA
      [Lactococcus lactis]"
      /codon_start=1
      /transl_table=11
      /product="conserved hypothetical protein"
      /protein_id="AA079754.1"
      /db_xref="GI:21904886"
      /translation="MKKSNLNNKWSFICLLAFNFAFLMVVASRLIQVREPESELIA
      KDPVKNKIGFTVITREQLNETVASYLKDYQTKKMSYKFTYATSSILFECTYQLQGYE
      IVYNLQIQNDKAVQLQVIFSFGVTLPLPEKDVLOLKKSYKLPSFVKVMPNQSA
      complement(645..1487)
      /gene="SpyM3_1148"
      complement(645..1487)

```

```

/gene="SpyM3_1148"
/notes="best blastp hit: emb|CAA72096.1| (v11213)
hypothetical protein [Streptococcus thermophilus]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA079755.1"
/db_xref="GI:21904887"
/translation="MNNRHLFSGIFPFVLSLCIAFLLLNIIIPKNSRLKSKDFLKKK
QVAIQVAIGDSLTGEGVGLTHQGGFVPLLTINDLSEYFRKANNVHONGYVSGDTSQIL
DRMKQKQLSLKRAIDMTLVGGNDVMAVIRKNLADLQVSFRPAKQIKSKTLKQI
IELARKNDKLPFIILGYLPNPFELTDMOKVIDDANTKTKEVGYDRYFVPI
NDLLYKINGIQEGIVHSSGDTTIVNDALFTGDFHPNNTGTQIMSNVMEKIKKHEK
KIP"
complement(1480..2319)
/gene="SpyM3_1149"
complement(1480..2319)
/gene="SpyM3_1149"
/notes="best non-GAS blastp hit: gb|AAK75223.1| (AE007412)
degV family protein [Streptococcus pneumoniae TIGR4]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA079756.1"
/db_xref="GI:21904888"
/translation="MGTIKIVTDSITIEPELIKALDITVPLVMSIDSKLYSDNDLK
EEGHFLSLMKASKSLPKTSOPPVGLFAETIENLVKGVTDIVAIHLSPALSGTIEASR
OGAEIAETPVTVLDSGTQDMKFOVVEAKAKAGASINELIAAQAIAKSKTELIG
VSTLENLVKGGRIQVTVGLSSLLNVKVMALKNDKELTLVKGKRTKWLDSVLA
KNSHRPIAIAISYAGEASLALFLKERIAAYNHSISVLETSGSIITHTYHGEGAFVAV
RYE"
complement(2547..3491)
/gene="SpyM3_1150"
complement(2547..3491)
/gene="SpyM3_1150"
/notes="best blastp hit: gb|AAL98083.1| (AE010066)
hypothetical protein [Streptococcus pyogenes MGAS8232]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA079757.1"
/db_xref="GI:21904889"
/translation="MTSKRACLSSIIVLASLTCGNQTVTSANHLSATGDKFDDCGSTLVE
KIVAPKDELEMLAWSSQTDDADRYEDFLDDDSFNDDSFNDSFISQNETDKMFEN
LTDRLDLEDEDTIEPQNVIMPDDDELDTDAVETRLTVSSAPHLAEALPKPHL
RSLSDTALRSGETIRGHLNDKLSVATKALTAQKFDLTTHVYSIGESFEVIAA
HYEDKRAESAFAKSKRHLPIATPDVVIEELRLVSSIGSKEDVSPVYNRKLGMAVA
KRKIALPQTGESESYYPALLGLMLGLTPIPIPKINN"
complement(3663..5324)
/gene="recl"
/notes="SpyM3_1151"
complement(3663..5324)
/gene="recl"
/notes="best blastp hit: gb|AAL98084.1| (AE010066) putative
DNA repair and genetic recombination protein
[Streptococcus pyogenes MGAS8232]"
/codon_start=1
/transl_table=11
/product="putative DNA repair and genetic recombination
protein"
/protein_id="AA079758.1"
/db_xref="GI:21904890"
/translation="MLLEISIKNFALIDEISLNFENGMTVLTGETGAGKSIIDAMNM
MLCARASTEVIRRGANKAETEGFVSDATPELVACLESSEGIAMEEBELIIRRDIFANGR
SVSRINGQMVNLTAKQVGLVDIHGHDQPELMRPHQOILDAFGKAPALAKEN
YQLIDPRYSLRQVIDKQNEKEKRDMDLAFQIAEIAAALSAGEDDRLNQERDR
LMEVSKQSDTIDQLDFDGRQIETFRDLTNSLTRYGGNVDVLDYDNIKEY
LLEEVSGKSLDLESELKALEKQVAAATNLSRRHQLAQLEAEIKAEKLYEMKA
DKVHFTTSKFNDRGNESLEFYISTNPGEKPLVKVHQAQLEAEIKAEKLYEMKA
DKTSIVDEVDTCVSGRVAQAIQAKIYKICRGHGVIAISHLPOVIALADYQYISKE
KEESTVSKVRLTPEERVEIASMIAGTDMTQAALTQARELLAKH"
complement(5346..5816)

```

gene

```

/codon_start=1
/transl_table=11
/product="glucose kinase"
/protein_id="AAL98113.1"
/db_xref="GI:19748683"
/transl_table="MSOKLIGIDLGTTIKFGILTAAGEVOEKWAIETNILEGGKHIV
PDIVASIKHRLDLYGLSADPVGIGMGSPGAVDRDNTVTGAFNLNWKETQEVGSVVE
KEICIPPAIDANDANVAALGERWVGAGENNPDVFMFTLGTGVGGGIIADGNLIHGVAGA
GGEGHMIPEPENGFACTCGSHCLETVASATGVVKVARLLAEAYEGDSAIKKAIDNG
EYTSKDIFMAAEGDSFADSVVEKVGYYILGLASANISNLINPDSVYVIGGVSAAGEF
LRSRIKFVTFPPQVRYSTKIKIAELGNDAGIIGAASLARQFTR"
/complement(3981..4184)
/gene="spym18.1547"
/complement(3981..4184)
/gene="spym18.1547"
/notes="best blastp match gb|AAK34324.1| (AE006586)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL98114.1"
/db_xref="GI:19748684"
/transl_table="MKTLYDVQQLKNFGIFVYLGKRLYDIEMMKIELQRLYDSGLLD
KRYLNAELILRREHRLLEKEG"
/complement(4326..4853)
/gene="spym18.1548"
/complement(4326..4853)
/gene="spym18.1548"
/notes="best blastp match gb|AAK34325.1| (AE006586)
putative peroxide resistance protein [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative peroxide resistance protein"
/protein_id="AAL98115.1"
/db_xref="GI:19748685"
/transl_table="MTNLTVENIVASVTHNISKKEASNEKTRAVLNQAVADLSVAAS
IVHQVHYMGPGLFLYLPKMDLSDLSNANLDMERLITIGCAPYSTLAEFSSKSHK
LDBAKGYDKTVAQHARLAEVYLYLSSLLQVGLDITDEBGADGNDLFTAAKTEAK
TIWMLQERQGGPAL"
/5036..5706
/gene="spym18.1549"
/5036..5706
/gene="spym18.1549"
/notes="best blastp match gb|AAK34326.1| (AE006586)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL98116.1"
/db_xref="GI:19748686"
/transl_table="MISFRINKYDVSFISCSGILLTHLFVCLLLYFPDNGQEQLSKQT
VIVCDQUTFAKRVFQNFITISOKIFFSLEISLIIILWASVHWVWSYCYLLILFS
LFLSFLDWRQSEYPTLIWLFPSVSLLEFYSINYLSLILGLLGLLAHLRPFISGAGDF
YLASLAIVLDLTSILNIIQSLASLAGITACULLGKIRIPFIPYLSFGLFWIVLLEHF"
/complement(5788..6867)
/gene="spym18.1550"
/complement(5788..6867)
/gene="spym18.1550"
/notes="best blastp match gb|AAK34327.1| (AE006586)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL98117.1"
/db_xref="GI:19748687"
/transl_table="MKPSIYSLTRDELIANAVERGOKQPRATQIWDWLYKKRVQSEFE
MTNISRDFVILNDSFCVNLKQRVQESADGVKYLFLPDGMLLETVLMRQHYGHS
VCVITQVCGNICGTFCASGLIKKORVQESADGVKYLFLPDGMLLETVLMRQHYGHS
GIEPFDNYKNVCMFLRVIINDNGLAIGARHITVSTVSGLAHKIRDFANEGVQVNLAYS
LHAPNDLRSSIMRVNRSFPLEKLFSAIEYYIEKTNRRVTFEYIMLNEVNDISIKQAE

```

gene

CDS

```

LADLTKTIRKLSYVNLIPYNPVSEHDOYSRSPKERVLAFLVDLVKKNGVNCVVRQEHCT
DIDAACGOLRSKTMKKDREKVATK"
/complement(6871..7515)
/gene="spym18.1551"
/complement(6871..7515)
/gene="spym18.1551"
/notes="best blastp match gb|AAK34328.1| (AE006586)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL98118.1"
/db_xref="GI:19748688"
/transl_table="MDFSERCOMKKEISPEWYNKPPGPKFTHFEQVQVKAEGIDLLI
LEVDKNAPDTSFGQRYTEVLLKYDIYVGDWNEQLRLKGFYKSDSDDKKTNRLSRLE
DYIKFCNFCATFVLENLPQDIKFEERQPKRKSXPKSRKSNRRKRNQQPATPK
SKKRVSKERQENQAPTQSKRRSNTHKKEKSRNQTSQLNTKISHFIIRKKDR"
/complement(7939..8916)
/gene="spym18.1552"
/complement(7939..8916)
/gene="spym18.1552"
/notes="best blastp match gb|AAK34329.1| (AE006586)
putative ribose transport operon repressor [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative ribose transport operon repressor"
/protein_id="AAL98119.1"
/db_xref="GI:19748689"
/transl_table="MTVIKQVAAEAGVSRTSVRYISQKGYVSDARHKIKAAIAKILH
LEEVYVHQLSKTNNAGIITTHDFTKRYPTLAIPVVYVDRVDQETQYGVFSDNRAGGL
AAQTVQACAKAEVLLIRGLDNAENINERFEASFYLQKODVTMYVCDSONDFESIQ
LEASYNLKCYPTIDSIAPSDIHAIYIHELHSGQKKIPQDVQIIGYDDILMSQFIYP

```

## Alignment Scores:

Pred. No.:	1.05e-42	Length:	11465
Score:	557.50	Matches:	107
Percent Similarity:	71.98%	Conservative:	42
Best Local Similarity:	51.69%	Mismatches:	41
Query Match:	55.25%	Indels:	17
DB:	1	Gaps:	4

US-09-868-352-23 (1-192) x AE010069 (1-11465)

Qy	1	MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe	20
Db	7491	ATGAAAAAGAAATTCGCCAGAAATGTACAACTATAATAAATTTCCAGGTCCTCCAAATTT	7432
Qy	21	IleHisPheGluAsnIleValLysSerAspPheGluPheGlnLeuValIleAsnGlu	40
Db	7431	ATTCATTTTGAAGACCAAGTTAAGCTGAAGCATTCATTTGTTACTCTTAGAAGATGTT	7372
Qy	41	LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyr	60
Db	7371	AAGAAGCGTTTGTATACGACTAGCTTTGGTCAACGTTATACAGAAGTCTCCTAAAGTAT	7312
Qy	61	AspPheIleValGlyAspTrpGlyAsnGluIleLeuArgLeuArgGlyPheTyrLysAsp	80
Db	7311	GACTATATTGTTGGCGATTGGGAAATGAACAGCTTCGCTTAAGAGGTTTATAAGGAT	7252
Qy	81	AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr	100
Db	7251	AGTGATCATATCAAGAAACCAATCGCATCTCAGCTTTAGAGGATTATATTAAGAAATTT	7192
Qy	101	CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe	120
Db	7191	TGCAATTTTGGTCTGCTTATTTTGTAGAAAAATCTTCACACAGATATTAAATTT	7132
Qy	121	AspAspGluArgProHisLysArgLysSer---ArgSerLysSerGln-----	136
Db	7131	GAAGAGGAGCCCAACCAAGACGAAAGAAATCACCTAAATCAAATCAATCGTCGCAAG	7072



Percent Similarity:	71.98%	Conservative:	42
Best Local Similarity:	51.69%	Mismatches:	41
Query Match:	55.35%	Indels:	17
DB:	1	Gaps:	4
US-09-868-352-23 (1-192) x AE006586 (1-11507)			
Qy	1	MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe	20
Db	7533	ATGAAAAAGAAATTTGCGCAGAAATGTACAACATATAAATTTTCCAGGTCCCAATTT	7474
Qy	21	IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu	40
Db	7473	ATTCATTTTGAAGACGAGTTAAGGCTGAAGGCATTTGATTCTCTAGAGATGTT	7414
Qy	41	LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyr	60
Db	7413	AAGAAGCGCTTTGATACGACTAGCTTTGTCACAGTTATACAGAAAGTTCTCCTAAAGTAT	7354
Qy	61	AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp	80
Db	7353	GACTATATTGTCGATTGGGGAATGAACAGCTTCGCTTAAAGGGTTTATAAGGAT	7294
Qy	81	AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr	100
Db	7293	AGTGACGATATCAAGAAACGAATCCATCTCACGTTTGAAGATTATATTAAGAATTT	7234
Qy	101	CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe	120
Db	7233	TGCAATTTTGGTTGCTTATTGTTTGTAGAAATCTTCATCCACAGATATTAAATTT	7174
Qy	121	AspAspGluArgProHisLysArgArgLysSer---ArgSerLysSerGln-----	136
Db	7173	GAAGAGGAGCGCCACCAACGACGAAAGAAATCACCTTAATCAAAATCAATCGTCGAAG	7114
Qy	137	-----SerSerLysSerGlnThrArgAsnAsn	145
Db	7113	CCAACTATTCAAATCAGCAGCGCCGACCACTAAGACGAATTCGAAG---CGGGCATCA	7057
Qy	146	ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg	165
Db	7056	AAAGAAAGCAACCTGAAAACCAAGCGTTTACCAGTCAAAACGTCGAAGTAACTAAG	6997
Qy	166	ArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSerAlaLysGlnHis	185
Db	6996	CATAAAGAAAGTCA---AAACGTAATCAGACTAGTCAACTTAATACCAAAATTAGTCAT	6940
Qy	186	LeuLeuPheValArgLysAsn	192
Db	6939	TTTATCATCAGAAAGAAGAT	6919
RESULT 3			
AE010069/c			
LOCUS	AE010069	11465 bp	DNA linear BCT 03-APR-2002
DEFINITION	Streptococcus pyogenes strain MGAS8232, section 117 of 173 of the complete genome.		
ACCESSION	AE010069	AE009949	
VERSION	AE010069.1	GI:19748679	
KEYWORDS			
SOURCE	Streptococcus pyogenes MGAS8232.		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
REFERENCE	1 (bases 1 to 11465)		
AUTHORS	Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M., Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M., Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M., Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.		
TITLE	Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)		
MEDLINE	21927593		

```

/db_xref="GI:13622615"
/translation="MTNLRNDIRNVAIIAHVDHCKTTLVDELLKQSHTLDERKLEQER
AMPSNDLEKRGITILAKNTAVAYNDVRINIMDTPGHADEFGEVERIMKWMDGVVLV
DAYEGTTPOTREVLKALKLQNLIPIVNVKIDKPSARPAEYVDEVLELELFCADDEQ
LEPVPYAGAINSLSDDPADQHTMAPIDFTIIDHIPAPVDNSDEPLQFOVSLDD
YNDVFGIRGVRFTVKVGDQVTLKSLDGTTKNFRVTKLFGFFGLERRETOEAKAG
DIATVAGMEDIFVGETITPTDCVEALPIILRIDETPLQMTFLVNSPFAEGEKWITSR
KVERLAELOTVDSRVDPDPSDKWTYSGRCELHLSILIEPMRREGYELQVSRPEV
LKEIDVSCPEPERVIDTPEYOCALIOSLSERKGDMLDMQMGVGTORLILFLIPA
RLGLIGSTELSTRGYIMNHTFDQYLPVVOGEIGRURGALVSYENKATYISIMR
IEERGIFVNGTVEYEGMIVGNSDNDLVNITAKQMTNVRSA TKQDTAVIKTPR
ITLESLEFLNDEYMEVTPESIRKURKQILNKAARDKANKKKKSAE"
gene
CDS
/notes="Best Blast hit = dbj|BAA12549.1| (D84432)
YqH [Bacillus subtilis] >gi|1303893|dbj|BAA12549.1| (D84432)
YqH [Bacillus subtilis] >gi|2634888|emb|CAB14385.1|
(299116) similar to hypothetical proteins [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34322.1"
/db_xref="GI:13622616"
/translation="MSPIYLILWLLVIGVYTYWYFSEFRMAKQVDNETKDVWVRQ
GLQIDUREAFAFKTHILGARNFPAQOFDAAIKGLRKDPVLIYENRPOYRVPVAVKK
LKAGFEDYVLRKGDYWDGKVKQTT"
complement(3046..4017)
gene
CDS
/notes="SpY1529"
complement(3046..4017)
/genes="gick"
/EC_number="2.7.1.2"
/notes="Best Blast hit = pir|T44787 glucokinase (EC
2.7.1.2) [imported] - Bacillus megaterium
>gi|2274967|emb|CAA03848.1| (AJ000005) glucose kinase
[Bacillus megaterium]"
/codon_start=1
/transl_table=11
/product="glucose kinase"
/protein_id="AAK34323.1"
/db_xref="GI:13622617"
/translation="MSQKLLGIDLGTTIKFGLTAAAGEVOEKWAIETNILEGKHIY
PDIIASIKRHLDDYGLUSSADFGIGMSPGAVDRDNTVTGAFNLWKKETQEVGSVVE
KELGTPAIDNDNAALGERWVAGENPNPVVFMTLGTGVGGIITADGNLHGVAGA
GGEIGHMIVEPENGFACTCGSHGCLTASATGVVKVARLLAEAYEGDSAIKAAIDNG
EGVTSKDIEMAAEAGDSFADSVVEKVGYYGLASANTSNLNPDSVVIIGGVSAAGEF
LRSRIEYFVTFPQVRYSTKIKIAELNDAGIIGAAASLARQFITK"
complement(4022..4225)
gene
CDS
/genes="SpY1530"
complement(4022..4225)
/notes="Best Blast hit = emb|CAA03847.1| (AJ000005) orf1
[Bacillus megaterium]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34324.1"
/db_xref="GI:13622618"
/translation="MKTLYDVQOLLKKNFGIFVYLGKRLYDIEMMKIELQRLYDSGLLD
KRDYLNALILREHRELEKEG"
complement(4367..4894)
gene
CDS
/genes="dpr"
complement(4367..4894)
/notes="Best Blast hit = gb|AAG33871.1|AF319974_1
(AF319974) dpr [Streptococcus suis]"
/codon_start=1

```

```

/transl_table=11
/product="putative peroxide resistance protein"
/protein_id="AAK34325.1"
/db_xref="GI:13622619"
/translation="MTNLTVENIYASVTHNISKEASKEKTKAVLNQAVADUSVAAS
IVHQHWYMRGFGFLYHPKMDLDSLANLDEMGERLITIGAPYSTLAEFKHSK
LOEAKTYOKTVAQHLARLVEVYLYSSLYQVGLDITDEEGDAGTNDLFTAAKTEAK
TIWMLQAEKQGGPAL"
5122..5748
gene
CDS
/genes="SpY1532"
5122..5748
/notes="Best Blast hit = dbj|BAA96471.1| (AB036428) type
IV prepilin peptidase homologue [Streptococcus mutans]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK34326.1"
/db_xref="GI:13622620"
/translation="MIAFFSICSGILLTHLYCLLLYFPDNGOQLSKOTVIYCDOLT
FAKRVPRQNFHITTSOKIFFSLEIGLIIILWASVMWVSASVCYLLLSLFLSFDW
RQCEYPPILWFSVSLLEFYSINYLISLILLGLLAHLRPFISIGADDFYLASALV
LDLTSULWLIQASLAGITACLLGLKIRPIPIYLSFGFLWIVLLEHF"
complement(5830..6909)
gene
CDS
/genes="SpY1533"
complement(5830..6909)
/notes="Best Blast hit = spj034617|YLON_BACSU HYPOTHETICAL
41.6 KDA PROTEIN IN FWT-SPOVM INTERGENIC REGION
>gi|7448849|pir|F69878 conserved hypothetical protein
YLON - Bacillus subtilis >gi|2337804|emb|CAA74265.1|
(Y13937) YLON protein [Bacillus subtilis]
>gi|2633947|emb|CAB13448.1| (299112) similar to
hypothetical proteins [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34327.1"
/db_xref="GI:13622621"
/translation="MKPSIYSLTRDELIAWVERGOKQFRATQIWDWLYKKRVQSFE
MNISQDFVILNDSFCVAPLQKRVVOESADGTVKYLFELPDGMLLETVMROHYGHS
VCVTTQVGCNIGTCFASGLIKKQDLNDSGETAQIMLVOKYFDDKQGERVSHVVM
GICEPFDNYKNMCFVRINDDGLAIGAEHITVSTGLAHKIRDFANEQGVQVNLAVS
LHAPNNDLRNSIMRVNRSFLEKLFSAIEYVTEKTRRVTFEYIMLNEVNDSTKQAE
LADLTKTIRKLSVNLPIPNVSEHDOYSRSPKRVLFADFVLKKNVCVVRQEHGT
DIDAGQLRSKTMKDKREKVTATK"
complement(6913..7557)
gene
CDS
/genes="SpY1534"
complement(6913..7557)
/notes="Best Blast hit = pir|F70023 hypothetical protein
yutD - Bacillus subtilis >gi|2635728|emb|CAB15221.1|
(299120) yutD [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK34328.1"
/db_xref="GI:13622622"
/translation="MDFSERQMKKEISPEMYNKNFPGPKFIHFPEQVKAEGIDLLL
LEDVKNAPDTSFGQRYTEVLLYDIYDVGWNEQURLKGYKSDSDIKRTNRISLE
DYIKFEFCNFCAYFVLENLHPQDIDKEEERQKPKSKSPKSNRRKPNYSNQQPATPK
SKSKRASKEQENQAFTSQKRRSNTKHKEKSKRNOTSOLNTKISHFIIRKKDK"
complement(7981..8958)
gene
CDS
/genes="SpY1535"
complement(7981..8958)
/notes="Best Blast hit = pir|I40462 repressor of ribose
transport operon - Bacillus subtilis
>gi|580917|emb|CAA81048.1| (225798) repressor of ribose

```

Alignment Scores:  
Pred. No.: 11507  
Score: 556.50  
Matches: 107

AX026665  
LOCUS AX026665 579 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 22 from Patent WO037646.  
ACCESSION AX026665  
VERSION AX026665.1 GI:10187839  
KEYWORDS Streptococcus sp. 'group B';  
SOURCE Streptococcus sp. 'group B';  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 579)  
AUTHORS Lane,J.D., Santangelo,J.D., Dougan,G., Everest,P., Feldman,R., Dobson,R.J., Hughes,M.J., Moore,J.C. and Wilson,R.K.  
TITLE Genes and proteins, and their use  
JOURNAL Patent: WO 0037646-A 22 29-JUN-2000;  
LANE JONATHAN DOUGLAS (GB); SANTANGELO JOSEPH DAVID (GB); DOUGAN GORDON (GB); EVEREST PAUL (GB); FELDMAN ROBERT (GB); DOBSON RICHARD JAMES (GB); HUGHES MARTIN JOHN GLENTON (GB); MICROSCIENCE LTD (GB); MOORE JOANNE CHRISTINE (GB); WILSON REBECCA KERRY (GB)  
FEATURES  
source  
1..579  
/organism="Streptococcus sp. 'group B'"  
/db\_xref="taxon:1319"  
CDS  
1..579  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC09097.1"  
/db\_xref="GI:10187840"  
translation="MRKEVTPMLNKNYPGQFIHFENIVKSDDEIFOLVINEKSAF  
DVTVGQRFSEILLKVDFTIVGDWNEQLRLGRFYKDSPIRKNSRLSLEDYIKVCN  
FCCAVFLNPDPRIKEDDERPHKRRKRSKSKSQTNRNRSQSNANAHFTSKRR  
KTKRQERHIKEQDQWNTSAKQLHLFVRKN"  
BASE COUNT 213 a 85 c 109 g 172 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,12e-87 Length: 579  
Score: 1009.00 Matches: 192  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
US-09-868-352-23 (1-192) x AX026665 (1-579)  
QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20  
Db 1 ATCGGAAAAGAGTGCACACAGAGATGCTTAACATAATAAGTATCTCGCCACAGTTT 60  
QY 21 IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40  
Db 61 ATTCACTTTGAAAATATCGTTAAAGATGATGATATTCAATTTTCAACTTGTATTATGAA 120  
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60  
Db 121 AAATCAGCTTTTCACTGACTGCTTTTGGACAACGTTTTCTGAGATTTATTAAATAT 180  
QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80  
Db 181 GATTTTATCGTTCGGCATTGGGGTACGAGCAGCTTGAGGCTAAGAGCITTTTACAAAGAT 240  
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100  
Db 241 GCTAGTACGATTAGAAAAATACCGCGATTTTCACGTTTAGAGATTATATTAAAGAGTAT 300  
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120  
Db 301 TGTAACTTTGGTTGCTTATTTTGGTTGGAGAATCCCAATCCTAGAGATATTAAATTT 360  
QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerLysSer 140  
Db 361 GATGATGAAGACCTCATAGCGCTCGTAGTCAAGATCCCAATCAACAATCATCAAGTCA 420

QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160  
Db 421 CAAACTAGAAATAATCGTCCCAAGTCAAATGCTCATTTTACAAAGTAAAGCGT 480  
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180  
Db 481 AAAGACACAAACGCCGTCGAAGAACGTCATATTAAAGAGAGCAAGATAAGGAATGACC 540  
QY 181 SerAlaLysGlnHisLeuLeuPheValArgLysAsn 192  
Db 541 TCTGCAACAGCAGATTGTATTTCGTAAGAAAAAT 576  
RESULT 2  
AE006586/11507 bp DNA linear BCT 01-JUN-2001  
LOCUS Streptococcus pyogenes M1 GAS strain SF370, section 115 of 167 of  
DEFINITION the complete genome.  
ACCESSION AE006586 AE004092  
VERSION AE006586.1 GI:13622613  
KEYWORDS Streptococcus pyogenes M1 GAS.  
SOURCE Streptococcus pyogenes M1 GAS  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 11507)  
AUTHORS Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K., Primeaux,C., Sezate,S., Surorov,A.N., Kenton,S., Lai,H., Lin,S., Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J., Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.  
TITLE Complete genome sequence of an M1 strain of Streptococcus pyogenes  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)  
MEDLINE 21192684  
PUBMED 11296296  
REFERENCE 2 (bases 1 to 11507)  
AUTHORS Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K., Primeaux,C., Sezate,S., Surorov,A.N., Kenton,S., Lai,H., Lin,S., Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J., Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.  
TITLE Direct Submission  
JOURNAL Submitted (10-APR-2001) Department of Microbiology and Immunology,  
University of Oklahoma Health Sciences Center, 940 SL Young Blvd,  
Oklahoma City, OK 73104, USA  
FEATURES  
Location/Qualifiers  
1..11507  
/organism="Streptococcus pyogenes M1 GAS"  
/strain="SF370"  
/serotype="M1"  
/db\_xref="taxon:160490"  
complement(250..501)  
/gene="SpY1526"  
complement(250..501)  
/gene="SpY1526"  
/note="Best Blastp hit = no blast hits"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK34320.1"  
/db\_xref="GI:13622614"  
translation="MFYLIMAVSIISYLYMAPKSVRNTLGMIGLVALLIVLAGL  
SFIKIMQTPPEFICMGVVALGYFALKDVRKTKPRVK"  
complement(623..2464)  
/gene="SpY1527"  
complement(623..2464)  
/gene="SpY1527"  
/note="Best Blastp hit = dbj|BAB06351.1| (AP001516)  
GTP-binding protein TypA/BipA (tyrosine phosphorylated  
protein A) (Bacillus halodurans]"  
/codon\_start=1  
/transl\_table=11  
/product="putative GTP-binding protein TypA/BipA (tyrosine  
phosphorylated protein A)"  
/protein\_id="AAK34321.1"

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 25, 2003, 09:38:47 ; Search time 2693 Seconds  
(without alignments)  
2074.912 Million cell updates/sec

Title: US-09-868-352-23  
Perfect score: 1009  
Sequence: 1 MRKEVTPEMLNKNYKPGQF.....EEQDKEMTSAKQHLFVRKN 192

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=fract+p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO\_spool/US09868352/runat\_14022003\_153516\_4220/app\_query.fasta\_1.391  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868352\_CGN\_1.1\_2425 @runat\_14022003\_153516\_4220 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_ov:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hun:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*

- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rdd:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1009	100.0	579	6	AX026665 Sequence
C 2	558.5	55.4	11507	1	AE006586 Streptoco
C 3	557.5	55.3	11465	1	AE010069 Streptoco
C 4	557.5	55.3	50463	1	AE014159 Streptoco
5	452.5	44.8	10029	1	AE008445 Streptoco
C 6	452.5	44.8	147006	2	SPNEU1904
7	438	43.4	5558	6	BD003783 Polynucle
C 8	438	43.4	10029	1	AE007384 Streptoco
9	331.5	32.9	10098	1	AE006255 Lactococc
C 10	255.5	25.3	3054	1	AF270167 Staphyloc
C 11	255.5	25.3	3054	1	AX145485 Sequence
12	249.5	24.7	290150	1	AP004824 Staphyloc
13	249.5	24.7	299050	1	AP003131 Staphyloc
14	249.5	24.7	348527	1	AP003360 Staphyloc
C 15	248	24.6	248050	1	AL596172 Listeria
C 16	248	24.6	349980	6	AX413018 Sequence
C 17	248	24.6	349980	6	AX417048 Sequence
18	243.5	24.1	390	6	AX142599 Sequence
C 19	240.5	23.7	324050	1	AL591983 Listeria
20	239.5	23.7	1234	6	AX414540 Sequence
21	239.5	23.7	2690	6	AX416642 Sequence
C 22	234.5	23.2	303250	1	AP001518 Bacillus
C 23	232.5	23.0	285	6	AX439521 Sequence
C 24	232.5	23.0	217420	1	BSUB0017 Bacillus su
25	222.5	22.1	273	6	AX434730 Sequence
26	113.5	11.2	6145	6	E26748 Structure a
27	113	11.2	179908	10	AC125344
28	109.5	10.9	4615	8	SCYGL133W
29	109.5	10.9	14607	8	SCX11
30	108.5	10.8	2204	10	BC019437
31	108.5	10.8	127044	2	AC098505 Rattus no
32	108	10.7	215455	2	AC083856 Mus muscu
33	107.5	10.7	3130	10	BC026944 Mus muscu
34	107.5	10.7	207814	10	AL590994 Mouse DNA
35	107	10.6	218677	2	AC109286 Mus muscu
36	106.5	10.6	25117	2	AC116102 Dictyoste
37	106.5	10.6	237855	2	AL391558 Homo sapi
C 38	104.5	10.4	100269	8	ATF18022 Arabidops
C 39	104.5	10.4	132934	2	AC120073 Rattus no
C 40	104.5	10.4	168111	2	AC096061 Rattus no
C 41	104	10.3	205764	2	AC079478 Mus muscu
C 42	103.5	10.3	191914	2	AC117251 Mus muscu
43	103.5	10.3	214403	2	AC126270 Mus muscu
44	102.5	10.2	69894	2	AC123604 Mus muscu
45	102.5	10.2	115049	2	AC073746 Mus muscu

ALIGNMENTS

RESULT 1

; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-09-984-198-150

Query Match 3.1%; Score 6; DB 10; Length 258;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTP 8  
Db 68 KEVTP 73

RESULT 40  
US-10-024-579-12  
; Sequence 12, Application US/10024579  
; Patent No. US20020119522A1  
; GENERAL INFORMATION:  
; APPLICANT: Friddle, Carl Johan  
; APPLICANT: Gerhardt, Brenda  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20020119522A1el Human Ion Channel-Related Proteins  
; FILE REFERENCE: LEX-0274-USA  
; CURRENT APPLICATION NUMBER: US/10/024,579  
; PRIOR FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: US 60/258,595  
; PRIOR FILING DATE: 2000-12-28  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-024-579-12

Query Match 3.1%; Score 6; DB 12; Length 264;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 SRISRL 93  
Db 80 SRISRL 85

Search completed: February 26, 2003, 10:05:16  
Job time : 20 secs

Db 54 SRISRL 59

## RESULT 37

US-10-024-579-14

; Sequence 14, Application US/10024579

; Patent No. US20020119522A1

; GENERAL INFORMATION:

; APPLICANT: Friddle, Carl Johan

; APPLICANT: Gerhardt, Brenda

; APPLICANT: Hilbun, Erin

; APPLICANT: Turner, C. Alexander Jr.

; TITLE OF INVENTION: No. US20020119522A1 Human Ion Channel-Related Proteins

; TITLE OF INVENTION: and Polynucleotides Encoding the Same

; FILE REFERENCE: LEX-0274-USA

; CURRENT APPLICATION NUMBER: US/10/024,579

; CURRENT FILING DATE: 2001-12-18

; PRIOR APPLICATION NUMBER: US 60/258,595

; PRIOR FILING DATE: 2000-12-28

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 14

; LENGTH: 257

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-024-579-14

Query Match 3.1%; Score 6; DB 12; Length 257;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 SRISRL 93

|||||

Db 54 SRISRL 59

## RESULT 38

US-09-752-639-150

; Sequence 150, Application US/09752639

; Patent No. US20020091243A1

; GENERAL INFORMATION:

; APPLICANT: Gatanaga, T.

; APPLICANT: Granger, G.A.

; TITLE OF INVENTION: Factors Altering Tumor Necrosis

; TITLE OF INVENTION: Factor Receptor Releasing Enzyme Activity, and Methods

; TITLE OF INVENTION: of Use Thereof

; NUMBER OF SEQUENCES: 154

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/752,639

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US99/10793

; FILING DATE:

; APPLICATION NUMBER: 09/081,385

; FILING DATE:

; APPLICATION NUMBER: 08/964,747

; FILING DATE: 05-NOV-1997

; APPLICATION NUMBER: 60/030,761

; FILING DATE: 06-NOV-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Wu, Frank

; REGISTRATION NUMBER: 41,386

; REFERENCE/DOCKET NUMBER: 22000-20577.21

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-813-5600

; TELEFAX: 650-494-0792

; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 150:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 258 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

US-09-752-639-150

Query Match 3.1%; Score 6; DB 10; Length 258;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTP 18

|||||

Db 68 KEVTP 173

## RESULT 39

US-09-984-198-150

; Sequence 150, Application US/09984198

; Patent No. US20020106679A1

; GENERAL INFORMATION:

; APPLICANT: Gatanaga, T.

; APPLICANT: Granger, G.A.

; TITLE OF INVENTION: Factors Altering Tumor Necrosis

; TITLE OF INVENTION: Factor Receptor Releasing Enzyme Activity, and Methods

; TITLE OF INVENTION: of Use Thereof

; NUMBER OF SEQUENCES: 154

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/984,198

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US99/10793

; FILING DATE:

; APPLICATION NUMBER: 09/081,385

; FILING DATE:

; APPLICATION NUMBER: 08/964,747

; FILING DATE: 05-NOV-1997

; APPLICATION NUMBER: 60/030,761

; FILING DATE: 06-NOV-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Wu, Frank

; REGISTRATION NUMBER: 41,386

; REFERENCE/DOCKET NUMBER: 22000-20577.21

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-813-5600

; TELEFAX: 650-494-0792

; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 150:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 258 amino acids

Db 138 VKSDDI 143

```
RESULT 33
US-09-774-639-155
; Sequence 155, Application US/09774639
; Publication No. US20030003555A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P1
; CURRENT APPLICATION NUMBER: US/09/774,639
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 155
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-774-639-155
```

```
Query Match 3.1%; Score 6; DB 9; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRK 134
|||||
Db 66 RKSRK 71
```

```
RESULT 34
US-09-738-626-4680
; Sequence 4680, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4680
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4680
```

```
Query Match 3.1%; Score 6; DB 9; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 SRISRL 93
|||||
Db 213 SRISRL 218
```

```
RESULT 35
US-09-738-626-6304
; Sequence 6304, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6304
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6304
```

```
Query Match 3.1%; Score 6; DB 9; Length 236;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 DKEMTS 181
|||||
Db 70 DKEMTS 75
```

```
RESULT 36
US-10-024-579-16
; Sequence 16, Application US/10024579
; Patent No. US20020119522A1
; GENERAL INFORMATION:
; APPLICANT: Fiddler, Carl Johan
; APPLICANT: Gerhardt, Brenda
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20020119522A1 Human Ion Channel-Related Proteins
; TITLE OF INVENTION: and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0274-USA
; CURRENT APPLICATION NUMBER: US/10/024,579
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US 60/258,595
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-024-579-16
```

```
Query Match 3.1%; Score 6; DB 12; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 SRISRL 93
|||||
```

; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA: 09/106,582  
; APPLICATION NUMBER: 09/106,582  
; FILING DATE: 29-JUN-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.439C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-622-4900  
; TELEFAX: 206-682-6031  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Ehrlichia  
US-09-159-469-32

Query Match 3.1%; Score 6; DB 10; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
|||||  
Db 138 VKSDDI 143

RESULT 30  
US-09-159-469-58  
; Sequence 58, Application US/09159469  
; Patent No. US20020064535A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
; TITLE OF INVENTION: THERAPY OF EHRlichia INFECTION  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/159,469  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/106,582  
; FILING DATE: 29-JUN-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.439C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-622-4900  
; TELEFAX: 206-682-6031  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 amino acids  
; TYPE: amino acid

; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-159-469-58

Query Match 3.1%; Score 6; DB 10; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
|||||  
Db 138 VKSDDI 143

RESULT 31  
US-09-798-042-32  
; Sequence 32, Application US/09798042  
; Patent No. US20020068343A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION  
; FILE REFERENCE: 210121.439C7  
; CURRENT APPLICATION NUMBER: US/09/798,042  
; CURRENT FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 32  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Ehrlichia sp.  
US-09-798-042-32

Query Match 3.1%; Score 6; DB 10; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
|||||  
Db 138 VKSDDI 143

RESULT 32  
US-09-798-042-58  
; Sequence 58, Application US/09798042  
; Patent No. US20020068343A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION  
; FILE REFERENCE: 210121.439C7  
; CURRENT APPLICATION NUMBER: US/09/798,042  
; CURRENT FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 58  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Ehrlichia sp.  
US-09-798-042-58

Query Match 3.1%; Score 6; DB 10; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
|||||



```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33652
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 285090.10
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 32
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1e+02
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.3e+02
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 56
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2e+02
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2e+02
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 26
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1e+02
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 90
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2e+02
; OTHER INFORMATION: EST_HUMAN HIT: BE742115.1, EVALUO 2.00e-49
; OTHER INFORMATION: EST_HUMAN HIT: BE299019.1, EVALUO 3.00e-49
; OTHER INFORMATION: SWISSPROT HIT: O43683, EVALUO 5.60e-01
; US-09-864-761-33652

Query Match
Best Local Similarity 100.0%; Pred. No. 1e+02; Length 138;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 DVTVEG 50
DB 43 DVTVEG 48
|||||
```

```

RESULT 27
US-09-764-864-1566
; Sequence 1566, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1566
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-1566

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Length 146;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 127 KRRKSR 132
DB 32 KRRKSR 37
|||||

RESULT 28
US-09-925-300-1731
; Sequence 1731, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1731
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (38)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (134)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (143)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-925-300-1731

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Length 156;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTPE 8
DB 106 KEVTPE 111
|||||

RESULT 29
US-09-159-469-32
; Sequence 32, Application US/09159469
; Patent No. US20020064535A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; THERAPY OF EHRlichia INFECTION
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/159,469
```

; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1527  
; LENGTH: 75  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-796-692-1527

Query Match 3.1%; Score 6; DB 9; Length 75;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 132 RSKSQS 137  
Db 67 RSKSQS 72  
|||||

RESULT 24  
US-09-764-869-945  
; Sequence 945, Application US/09764869  
; Patent No. US20020061521A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007  
; CURRENT APPLICATION NUMBER: US/09/764,869  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2442  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 945  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (2)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (29)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-869-945

Query Match 3.1%; Score 6; DB 10; Length 88;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 145 NRSQSN 150  
Db 75 NRSQSN 80  
|||||

RESULT 25  
US-09-925-299-1067  
; Sequence 1067, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1067  
; LENGTH: 98  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (4)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (8)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (27)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (37)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (68)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (69)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (73)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (83)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1067

Query Match 3.1%; Score 6; DB 10; Length 98;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 KEVTPE 8  
Db 75 KEVTPE 80  
|||||

RESULT 26  
US-09-864-761-33f52  
; Sequence 33652, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Pern, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Henzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Acomlca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669

GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aemica-x-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
SEQ ID NO 35132  
LENGTH: 49  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC009955.1  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9  
US-09-864-761-35132

Query Match 3.1%; Score 6; DB 10; Length 49;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KRQER 169  
|||||  
DB 24 KRQER 29

RESULT 22  
US-09-764-878-190  
Sequence 190, Application US/09764878  
Patent No. US20020090615A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PA121  
CURRENT APPLICATION NUMBER: US/09/764,878  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 428  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 190  
LENGTH: 73  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (34)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (35)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (70)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (73)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-878-190

Query Match 3.1%; Score 6; DB 10; Length 73;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 LEDYIK 98  
|||||  
DB 23 LEDYIK 28

RESULT 23  
US-09-796-692-1527  
Sequence 1527, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950

;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 40249  
;; LENGTH: 38  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC009155.3  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.3  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.8  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.7  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.7  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.5  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.5  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.1  
;; OTHER INFORMATION: EST\_HUMAN HIT: AUL33018.1, EVALUE 4.00e-05  
US-09-864-761-40249

Query Match 3.1%; Score 6; DB 10; Length 38;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SEILLK 59  
DB 29 SEILLK 34

RESULT 20  
US-09-864-761-37254  
; Sequence 37254, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

;; FILE REFERENCE: Aeomica-X-1  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 09/864,761  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 37251  
;; LENGTH: 42  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC007314.3  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.93  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.88  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.82  
;; OTHER INFORMATION: EST\_HUMAN HIT: BE184944.1, EVALUE 4.00e-16  
US-09-864-761-37254

Query Match 3.1%; Score 6; DB 10; Length 42;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 KRRKSR 132  
DB 2 KRRKSR 7

RESULT 21  
US-09-864-761-35132  
; Sequence 35132, Application US/09864761  
; Patent No. US20020048763A1

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138  
|||||  
Db 267 RSKSQSS 273

## RESULT 16

US-10-085-027-1  
; Sequence 1, Application US/10085027  
; Patent No. US20020132759A1  
; GENERAL INFORMATION:  
; APPLICANT: YAZAKI, YOSHIO  
; APPLICANT: ASANO, TOMOICHIRO  
; APPLICANT: KUBO, HIDEO  
; APPLICANT: KANDA, AKIRA  
; TITLE OF INVENTION: REMEDIES FOR DISEASES CAUSED BY INSULIN RESISTANCE  
; FILE REFERENCE: 4895-0019-0PCT  
; CURRENT APPLICATION NUMBER: US/10/085,027  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: PCT/JP98/04293  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: JP9-263719  
; PRIOR FILING DATE: 1997-09-29  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 1242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-085-027-1

Query Match 3.6%; Score 7; DB 12; Length 1242;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138  
|||||  
Db 267 RSKSQSS 273

## RESULT 17

US-10-038-612-27  
; Sequence 27, Application US/10038612  
; Patent No. US20020160478A1  
; GENERAL INFORMATION:  
; APPLICANT: Ben-Sasson, Shmuel A.  
; TITLE OF INVENTION: Short Peptides Which Selectively  
; TITLE OF INVENTION: Modulate the Activity of Protein Kinases  
; FILE REFERENCE: 1242.1029-000 (CMCC-679)  
; CURRENT APPLICATION NUMBER: US/10/038,612  
; CURRENT FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US 09/161,094  
; PRIOR FILING DATE: 1998-09-25  
; NUMBER OF SEQ ID NOS: 172  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 27  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: unknown  
; FEATURE:  
; OTHER INFORMATION: POLO  
US-10-038-612-27

Query Match 3.1%; Score 6; DB 9; Length 19;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 HKRRKS 131  
|||||  
Db 13 HKRRKS 18

## RESULT 18

US-09-071-838-217  
; Sequence 217, Application US/09071838  
; Patent No. US20020152501A1  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Ohad, Nir  
; APPLICANT: Kiyosue, Tomohiro  
; APPLICANT: Yadegari, Ramin  
; APPLICANT: Margossian, Linda  
; APPLICANT: Harada, John  
; APPLICANT: Goldberg, Robert B.  
; TITLE OF INVENTION: Nucleic Acids That Control Seed and  
; TITLE OF INVENTION: Fruit Development in Plants  
; NUMBER OF SEQUENCES: 324  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,838  
; FILING DATE: 01-MAY-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-0861000US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 217:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-071-838-217

Query Match 3.1%; Score 6; DB 10; Length 25;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 KNSRIS 91  
|||||  
Db 9 KNSRIS 14

## RESULT 19

US-09-864-761-40249  
; Sequence 40249, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Acomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04

; APPLICANT: Deutch, Alan H  
; APPLICANT: Ghanbari, Hossein A  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS  
; FILE REFERENCE: 21486-032 CIP  
; CURRENT APPLICATION NUMBER: US/09/859,604  
; CURRENT FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/436,184  
; PRIOR FILING DATE: 1999-11-08  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-859-604-5

Query Match 3.6%; Score 7; DB 10; Length 1242;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138  
| | | | |  
Db 267 RSKSQSS 273

RESULT 12  
US-09-903-063-5  
; Sequence 5, Application US/09903063  
; Patent No. US20020114810A1  
; GENERAL INFORMATION:  
; APPLICANT: Wands, Jack R.  
; APPLICANT: de la Monte, Suzanne M.  
; APPLICANT: Ince, Nedim  
; APPLICANT: Carlson, Rolf I.  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS  
; FILE REFERENCE: 21486-032 DIV3  
; CURRENT APPLICATION NUMBER: US/09/903,063  
; CURRENT FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: 09/436,184  
; PRIOR FILING DATE: 1999-11-08  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-903-063-5

Query Match 3.6%; Score 7; DB 10; Length 1242;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138  
| | | | |  
Db 267 RSKSQSS 273

RESULT 13  
US-09-903-216-5  
; Sequence 5, Application US/09903216  
; Patent No. US20020114811A1  
; GENERAL INFORMATION:  
; APPLICANT: Wands, Jack R.  
; APPLICANT: de la Monte, Suzanne M.  
; APPLICANT: Ince, Nedim  
; APPLICANT: Carlson, Rolf I.  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS  
; FILE REFERENCE: 21486-032 DIV2  
; CURRENT APPLICATION NUMBER: US/09/903,216  
; CURRENT FILING DATE: 2001-07-11  
; PRIOR APPLICATION NUMBER: 09/436,184  
; PRIOR FILING DATE: 1999-11-08  
; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-903-216-5

Query Match 3.6%; Score 7; DB 10; Length 1242;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138  
| | | | |  
Db 267 RSKSQSS 273

RESULT 14  
US-09-903-199-5  
; Sequence 5, Application US/09903199  
; Patent No. US20020122802A1  
; GENERAL INFORMATION:  
; APPLICANT: Wands, Jack R.  
; APPLICANT: de la Monte, Suzanne M.  
; APPLICANT: Ince, Nedim  
; APPLICANT: Carlson, Rolf I.  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS  
; FILE REFERENCE: 21486-032 DIV4  
; CURRENT APPLICATION NUMBER: US/09/903,199  
; CURRENT FILING DATE: 2001-07-11  
; PRIOR APPLICATION NUMBER: 09/436,184  
; PRIOR FILING DATE: 1999-11-08  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-903-199-5

Query Match 3.6%; Score 7; DB 10; Length 1242;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138  
| | | | |  
Db 267 RSKSQSS 273

RESULT 15  
US-09-903-023-5  
; Sequence 5, Application US/09903023  
; Patent No. US20020146421A1  
; GENERAL INFORMATION:  
; APPLICANT: Wands, Jack R.  
; APPLICANT: de la Monte, Suzanne M.  
; APPLICANT: Ince, Nedim  
; APPLICANT: Carlson, Rolf I.  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS  
; FILE REFERENCE: 21486-032 DIV1  
; CURRENT APPLICATION NUMBER: US/09/903,023  
; CURRENT FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: 09/436,184  
; PRIOR FILING DATE: 1999-11-08  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-903-023-5

Query Match 3.6%; Score 7; DB 10; Length 1242;  
Best Local Similarity 100.0%; Pred. No. 68;

NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/146.474  
FILING DATE: 14-May-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/937,067  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Lehnhardt, Susan K.  
REGISTRATION NUMBER: 33,943  
REFERENCE/DOCKET NUMBER: 23647-20018.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 813-5600  
TELEFAX: (650) 494-0792  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 317 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-146-474-6

Query Match 3.6%; Score 7; DB 9; Length 317;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 KRKDKR 165  
DB 236 KRKDKR 242

RESULT 8  
US-09-866-582-16  
; Sequence 16, Application US/09866582  
; Patent No. US20020127620A1  
; GENERAL INFORMATION:  
; APPLICANT: Witman, George B.  
; APPLICANT: Pazour, Gregory J.  
; APPLICANT: Rosenbaum, Joel L.  
; APPLICANT: Cole, Douglas G.  
; TITLE OF INVENTION: INTRACELLULAR TRANSPORT  
; FILE REFERENCE: 07917-145001  
; CURRENT APPLICATION NUMBER: US/09/866,582  
; CURRENT FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: US 60/206,923  
; PRIOR FILING DATE: 2000-05-24  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 782  
; TYPE: PRT  
; ORGANISM: Chlamydomonas reinhardtii  
US-09-866-582-16

Query Match 3.6%; Score 7; DB 10; Length 782;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 DWGNEQL 72

DB 769 DWGNEQL 775  
RESULT 9  
US-09-436-184-5  
; Sequence 5, Application US/09436184  
; Publication No. US20030031670A1  
; GENERAL INFORMATION:  
; APPLICANT: Wands, Jack R.  
; APPLICANT: de la Monte, Suzanne M.  
; APPLICANT: Ince, Nedim  
; APPLICANT: Carlson, Rolf I.  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS  
; FILE REFERENCE: R.I. Hosp. - Malignant Neoplasms  
; CURRENT APPLICATION NUMBER: US/09/436,184  
; CURRENT FILING DATE: 1999-11-08  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-436-184-5  
Query Match 3.6%; Score 7; DB 9; Length 1242;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138  
DB 267 RSKSQSS 273

RESULT 10  
US-09-903-248-5  
; Sequence 5, Application US/09903248  
; Patent No. US20020102283A1  
; GENERAL INFORMATION:  
; APPLICANT: Wands, Jack R.  
; APPLICANT: de la Monte, Suzanne M.  
; APPLICANT: Ince, Nedim  
; APPLICANT: Carlson, Rolf I.  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS  
; FILE REFERENCE: 21486-032 DIV5  
; CURRENT APPLICATION NUMBER: US/09/903,248  
; CURRENT FILING DATE: 2001-07-11  
; PRIOR APPLICATION NUMBER: 09/436,184  
; PRIOR FILING DATE: 1999-11-08  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-903-248-5

Query Match 3.6%; Score 7; DB 10; Length 1242;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138  
DB 267 RSKSQSS 273

RESULT 11  
US-09-859-604-5  
; Sequence 5, Application US/09859604  
; Patent No. US20020110559A1  
; GENERAL INFORMATION:  
; APPLICANT: Wands, Jack R.  
; APPLICANT: de la Monte, Suzanne M

; CURRENT APPLICATION NUMBER: US/10/085,027  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: 09/508,691  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: PCT/JP98/04293  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: JP9-263719  
; PRIOR FILING DATE: 1997-09-29  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Description of Artificial Sequence:synthetic peptide  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (9)..(9)  
; OTHER INFORMATION: PHOSPHORYLATION  
US-10-085-027-2

Query Match 3.6%; Score 7; DB 12; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RKSQSS 138  
|||||  
Db 6 RKSQSS 12

## RESULT 4

US-10-117-846-19  
; Sequence 19, Application US/10117846  
; Patent No. US20020168673A1  
; GENERAL INFORMATION:  
; APPLICANT: Fuller, Margaret T  
; APPLICANT: Hales, Karen G.  
; APPLICANT: Santel, Ansgar H.  
; TITLE OF INVENTION: Mitofusins, Fzo Homologs and Functional  
; TITLE OF INVENTION: Derivatives Thereof  
; FILE REFERENCE: STAN-063CIP3  
; CURRENT APPLICATION NUMBER: US/10/117,846  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: 09/413,285  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: PCT/US00/27871  
; PRIOR FILING DATE: 2000-10-06  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: S. cerevisiae  
US-10-117-846-19

Query Match 3.6%; Score 7; DB 9; Length 157;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 FSEILK 59  
|||||  
Db 53 FSEILK 59

## RESULT 5

US-10-085-027-5  
; Sequence 5, Application US/10085027  
; Patent No. US20020132759A1  
; GENERAL INFORMATION:  
; APPLICANT: YAZAKI, YOSHIO  
; APPLICANT: ASANO, TONOICHIRO

; APPLICANT: KUBO, HIDEO  
; APPLICANT: KANDA, AKIRA  
; TITLE OF INVENTION: REMEDIES FOR DISEASES CAUSED BY INSULIN RESISTANCE  
; FILE REFERENCE: 4895-0019-0PCT  
; CURRENT APPLICATION NUMBER: US/10/085,027  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: 09/508,691  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: PCT/JP98/04293  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: JP9-263719  
; PRIOR FILING DATE: 1997-09-29  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-085-027-5

Query Match 3.6%; Score 7; DB 12; Length 159;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RKSQSS 138  
|||||  
Db 108 RKSQSS 114

## RESULT 6

US-09-867-550-1934  
; Sequence 1934, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: Thereby  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1934  
; LENGTH: 200  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-867-550-1934

Query Match 3.6%; Score 7; DB 10; Length 200;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRSK 135  
|||||  
Db 42 RKSRSK 48

## RESULT 7

US-10-146-474-6  
; Sequence 6, Application US/10146474  
; Publication No. US20030023061A1  
; GENERAL INFORMATION:  
; APPLICANT: Umansky, Samuil  
; APPLICANT: Melkonyan, Hovsep  
; TITLE OF INVENTION: A FAMILY OF GENES ENCODING  
; APOPTOSIS-RELATED PEPTIDES; PEPTIDES ENCODED THEREBY AND  
; METHODS OF USE THEREOF



969 5 2.6 272 9 US-09-828-523A-2 Sequence 2, Appli  
970 5 2.6 272 9 US-09-738-626-5598 Sequence 5398, Ap  
971 5 2.6 274 10 US-09-815-242-12496 Sequence 12496, A  
972 5 2.6 275 10 US-09-051-755-5 Sequence 5, Appli  
973 5 2.6 276 9 US-09-738-626-4122 Sequence 4122, Ap  
974 5 2.6 276 10 US-09-864-761-46690 Sequence 46690, A  
975 5 2.6 277 10 US-09-923-300-1169 Sequence 1169, Ap  
976 5 2.6 277 10 US-09-923-300-1467 Sequence 1467, Ap  
977 5 2.6 278 10 US-09-810-264-28 Sequence 28, Appl  
978 5 2.6 278 10 US-09-246-129B-6 Sequence 6, Appli  
979 5 2.6 278 10 US-09-872-523-41 Sequence 41, Appl  
980 5 2.6 278 10 US-09-899-059-6 Sequence 6, Appli  
981 5 2.6 279 8 US-08-971-317A-4 Sequence 4, Appli  
982 5 2.6 279 9 US-10-017-910-5 Sequence 5, Appli  
983 5 2.6 279 10 US-09-864-761-34389 Sequence 34389, A  
984 5 2.6 279 10 US-09-193-663-4 Sequence 4, Appli  
985 5 2.6 279 10 US-09-861-451A-28 Sequence 28, Appl  
986 5 2.6 279 12 US-10-014-927-19 Sequence 19, Appl  
987 5 2.6 280 10 US-09-835-788A-16 Sequence 16, Appl  
988 5 2.6 281 8 US-08-971-317A-6 Sequence 6, Appli  
989 5 2.6 281 9 US-09-131-237-6 Sequence 6, Appli  
990 5 2.6 281 9 US-09-802-669-25 Sequence 25, Appl  
991 5 2.6 281 10 US-09-193-663-6 Sequence 6, Appli  
992 5 2.6 281 10 US-09-027-287-6 Sequence 6, Appli  
993 5 2.6 281 10 US-09-252-656B-6 Sequence 6, Appli  
994 5 2.6 281 10 US-09-929-493-6 Sequence 6, Appli  
995 5 2.6 281 10 US-09-927-110-1 Sequence 1, Appli  
996 5 2.6 281 12 US-10-012-452-13 Sequence 13, Appl  
997 5 2.6 282 9 US-10-167-555-2 Sequence 2, Appli  
998 5 2.6 282 10 US-09-925-300-1694 Sequence 1694, Ap  
999 5 2.6 283 10 US-09-925-302-533 Sequence 533, App  
1000 5 2.6 283 10 US-09-925-302-693 Sequence 693, App

ALIGNMENTS

RESULT 1  
US-09-815-242-13414  
; Sequence 13414, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13414  
; LENGTH: 176

; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13414  
Query Match 6.2%; Score 12; DB 10; Length 176;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 NEQLRLRGFYKD 80  
|||||  
DB 69 NEQLRLRGFYKD 80  
RESULT 2  
US-09-815-242-13567  
; Sequence 13567, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13567  
; LENGTH: 176  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13567  
Query Match 6.2%; Score 12; DB 10; Length 176;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 NEQLRLRGFYKD 80  
|||||  
DB 69 NEQLRLRGFYKD 80  
RESULT 3  
US-10-085-027-2  
; Sequence 2, Application US/10085027  
; Patent No. US20020132759A1  
; GENERAL INFORMATION:  
; APPLICANT: YAZAKI, YOSHIO  
; APPLICANT: ASANO, TOMOICHIRO  
; APPLICANT: KUBO, HIDEO  
; APPLICANT: KANDA, AKIRA  
; TITLE OF INVENTION: REMEDIES FOR DISEASES CAUSED BY INSULIN RESISTANCE  
; FILE REFERENCE: 4895-0019-0PCT















GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 10:01:14 ; Search time 14 Seconds  
(without alignments)  
517.328 Million cell updates

Title: US-09-868-352-23

Perfect score:

Sequence: 1 MRKEVTPEMLNKNYPGPQF.....EEQDKEMTSAQHLLFVRKN 192

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 174566 seqs, 37721826 residues

Word size :

Total number of hits satisfying chosen parameters: 174566

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

```

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	12	6.2	176	10	US-09-815-242-13414		Sequence 13414, A
2	12	6.2	176	10	US-09-815-242-13567		Sequence 13567, A
3	7	3.6	15	12	US-10-085-027-2		Sequence 2, Appli
4	7	3.6	157	9	US-10-117-846-19		Sequence 19, Appli
5	7	3.6	159	12	US-10-085-027-5		Sequence 5, Appli
6	7	3.6	200	10	US-09-867-550-1934		Sequence 1934, Ap
7	7	3.6	317	9	US-10-146-474-6		Sequence 6, Appli
8	7	3.6	782	10	US-09-866-582-16		Sequence 16, Appli
9	7	3.6	1242	9	US-09-436-184-5		Sequence 5, Appli
10	7	3.6	1242	10	US-09-903-248-5		Sequence 5, Appli
11	7	3.6	1242	10	US-09-859-604-5		Sequence 5, Appli
12	7	3.6	1242	10	US-09-903-063-5		Sequence 5, Appli
13	7	3.6	1242	10	US-09-903-216-5		Sequence 5, Appli
14	7	3.6	1242	10	US-09-903-199-5		Sequence 5, Appli
15	7	3.6	1242	10	US-09-903-023-5		Sequence 5, Appli
16	7	3.6	1242	12	US-10-085-027-1		Sequence 1, Appli
17	6	3.1	19	9	US-10-038-612-27		Sequence 27, Appli
18	6	3.1	25	10	US-09-071-838-217		Sequence 217, App
19	6	3.1	38	10	US-09-864-761-40249		Sequence 40249, A



TELEPHONE: 206-622-4900  
TELEFAX: 206-682-6031  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-106-582-58

Query Match 3.1%; Score 6; DB 4; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
Db 138 VKSDDI 143

RESULT 38  
US-09-134-001C-3459  
Sequence 3459, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3459  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3459

Query Match 3.1%; Score 6; DB 4; Length 198;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 FENIVK 28  
Db 25 FENIVK 30

RESULT 39  
US-09-134-001C-4784  
Sequence 4784, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4784  
LENGTH: 244  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4784

Query Match 3.1%; Score 6; DB 4; Length 244;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 SOSNAN 152  
Db 229 SOSNAN 234

RESULT 40  
US-08-252-995D-11  
Sequence 11, Application US/08252995D  
Patent No. 5650501  
GENERAL INFORMATION:  
APPLICANT: Dennis, James W  
APPLICANT: Heffernan, Mike  
APPLICANT: Fode, Carol  
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/252,995D  
FILING DATE: 02-JUN-1994  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurdydyk, Linda M  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 271 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Drosophila melanogaster  
US-08-252-995D-11

Query Match 3.1%; Score 6; DB 1; Length 271;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 HKRRKS 131  
Db 96 HKRRKS 101

Search completed: February 26, 2003, 10:02:05  
Job time : 31 secs

RESULT 34  
US-09-295-028-32  
; Sequence 32, Application US/09295028  
; Patent No. 6277381  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION  
; FILE REFERENCE: 210121.439C4  
; CURRENT APPLICATION NUMBER: US/09/295,028  
; CURRENT FILING DATE: 1999-04-20  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 32  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Ehrlichia sp.  
US-09-295-028-32

Query Match 3.1%; Score 6; DB 4; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
|||||

Db 138 VKSDDI 143

RESULT 35  
US-09-295-028-58  
; Sequence 58, Application US/09295028  
; Patent No. 6277381  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION  
; FILE REFERENCE: 210121.439C4  
; CURRENT APPLICATION NUMBER: US/09/295,028  
; CURRENT FILING DATE: 1999-04-20  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 58  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Ehrlichia sp.  
US-09-295-028-58

Query Match 3.1%; Score 6; DB 4; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
|||||

Db 138 VKSDDI 143

RESULT 36  
US-09-106-582-32  
; Sequence 32, Application US/09106582  
; Patent No. 6306402  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; THERAPY

; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/106,582  
; FILING DATE: 29-JUN-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.439C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-622-4900  
; TELEFAX: 206-682-6031  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Ehrlichia  
; US-09-106-582-32

Query Match 3.1%; Score 6; DB 4; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
|||||

Db 138 VKSDDI 143

RESULT 37  
US-09-106-582-58  
; Sequence 58, Application US/09106582  
; Patent No. 6306402  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/106,582  
; FILING DATE: 29-JUN-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.439C2  
; TELECOMMUNICATION INFORMATION:

ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/975,762  
APPLICATION NUMBER: US/08/975,762  
FILING DATE: 21-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.439  
TELEPHONE: 206-622-4900  
TELEFAX: 206-622-4900  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE: Ehrlichia  
ORGANISM: Ehrlichia  
US-08-975-762-32

Query Match 3.1%; Score 6; DB 4; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
Db 138 VKSDDI 143

RESULT 32  
US-08-975-762-58  
Sequence 58, Application US/08975762  
Patent No. 6207169  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,762  
FILING DATE: 21-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.439  
TELECOMMUNICATION INFORMATION:

## TREATMEN

TELEPHONE: 206-622-4900  
TELEFAX: 206-682-6031  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-975-762-58

Query Match 3.1%; Score 6; DB 4; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
Db 138 VKSDDI 143

RESULT 33  
US-08-821-324-32  
Sequence 32, Application US/08821324  
Patent No. 6231869  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/821,324  
APPLICATION NUMBER: US/08/821,324  
FILING DATE: 21-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.439  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900  
TELEFAX: 206-682-6031  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE: Ehrlichia  
ORGANISM: Ehrlichia  
US-08-821-324-32

## THERAP

Query Match 3.1%; Score 6; DB 4; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32  
Db 138 VKSDDI 143

ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,010A  
FILING DATE: February 15, 1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/863-9700  
TELEFAX: 303/863-0223  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 149 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-602-010A-18

Query Match 3.1%; Score 6; DB 1; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRK 161  
|||||  
Db 109 TSKRRK 114

## RESULT 29

US-08-680-726A-18  
Sequence 18, Application US/08680726A  
Patent No. 5804197  
GENERAL INFORMATION:  
APPLICANT: Haanes, Elizabeth J.  
APPLICANT: Frank, Rexann S.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680,726A  
FILING DATE: 12-JUL-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 149 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-680-726A-18

Query Match 3.1%; Score 6; DB 1; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRK 161  
|||||  
Db 109 TSKRRK 114

## RESULT 30

US-09-092-409-18  
Sequence 18, Application US/09092409  
Patent No. 6151478  
GENERAL INFORMATION:  
APPLICANT: Haanes, Elizabeth J.  
APPLICANT: Frank, Rexann S.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/092,409  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/680,726  
FILING DATE: 12-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 149 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-092-409-18

Query Match 3.1%; Score 6; DB 4; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRK 161  
|||||  
Db 109 TSKRRK 114

## RESULT 31

US-08-975-762-35  
Sequence 32, Application US/08975762  
Patent No. 6207169  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
CORRESPONDENCE ADDRESS:  
NUMBER OF SEQUENCES: 73

## US-08-680-726A-18

APPLICATION NUMBER: US/09/092,409  
FILING DATE: 12-JUL-1996  
CLASSIFICATION: 3.1%  
PRIOR APPLICATION NUMBER: 08/680,726  
FILING DATE: 12-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-092-409-62

Query Match 3.1%  
Best Local Similarity 100.0%  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 TSKRRK 161  
Db 79 TSKRRK 84

RESULT 26  
US-08-680-726A-66  
Sequence 66, Application US/08680726A  
Patent No. 5804197  
GENERAL INFORMATION:  
APPLICANT: Haanes, Elizabeth J.  
APPLICANT: Frank, Rexann S.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680,726A  
FILING DATE: 12-JUL-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 139 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-680-726A-66

Query Match 3.1%  
Best Local Similarity 100.0%  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQS 137  
Db 120 RSKSQS 125

RESULT 27  
US-09-092-409-66  
Sequence 66, Application US/09092409  
Patent No. 6159478  
GENERAL INFORMATION:  
APPLICANT: Haanes, Elizabeth J.  
APPLICANT: Frank, Rexann S.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/092,409  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/680,726  
FILING DATE: 12-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 139 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-092-409-66

Query Match 3.1%  
Best Local Similarity 100.0%  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQS 137  
Db 120 RSKSQS 125

RESULT 28  
US-08-602-010A-18  
Sequence 18, Application US/08602010A  
Patent No. 5753235  
GENERAL INFORMATION:  
APPLICANT: Haanes, Elizabeth J.  
APPLICANT: Frank, Rexann S.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.

Query Match 3.1%  
Best Local Similarity 100.0%  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454312-2074  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 108 amino acids  
TYPE: amino acid  
STRANDEDNESS: n/a  
TOPOLOGY: linear  
MOLECULE TYPE: amino acid  
US-08-710-749-25

Query Match 3.1%; Score 6; DB 2; Length 108;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 EDYKE 99  
|||||  
Db 68 EDYKE 73

## RESULT 23

US-08-710-749-26  
Sequence 26, Application US/08710749  
Patent No. 5955089

GENERAL INFORMATION:  
APPLICANT: Briles, David E.  
APPLICANT: Hollingshead, Susan  
APPLICANT: Becker, Robert  
TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE  
TITLE OF INVENTION: PROTEINS  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/710,749  
FILING DATE: 20-SEP-1996  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454312-2074  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 108 amino acids  
TYPE: amino acid  
STRANDEDNESS: n/a  
TOPOLOGY: linear  
MOLECULE TYPE: amino acid  
US-08-710-749-26

Query Match 3.1%; Score 6; DB 2; Length 108;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 EDYKE 99  
|||||  
Db 68 EDYKE 73

## RESULT 24

US-08-680-726A-6;  
Sequence 62, Application US/08680726A  
Patent No. 5804197  
GENERAL INFORMATION:  
APPLICANT: Haanes, Elizabeth J.  
APPLICANT: Frank, Rexann S.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680,726A  
FILING DATE: 12-JUL-1996  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-680-726A-62

Query Match 3.1%; Score 6; DB 1; Length 119;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRK 161  
|||||  
Db 79 TSKRRK 84

## RESULT 25

US-09-092-409-6;  
Sequence 62, Application US/09092409  
Patent No. 6159478  
GENERAL INFORMATION:  
APPLICANT: Haanes, Elizabeth J.  
APPLICANT: Frank, Rexann S.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-710-749-22

Query Match
Best Local Similarity 100.0%; DB 2; Length 108;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 EDYKE 99
Db 68 EDYKE 73

RESULT 20
US-08-710-749-23
; Sequence 23, Application US/08710749
; Patent No. 5955089
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Becker, Robert
; TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,749
; FILING DATE: 20-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2074
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: n/a
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-710-749-24

Query Match
Best Local Similarity 100.0%; DB 2; Length 108;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 EDYKE 99
Db 68 EDYKE 73

RESULT 22
US-08-710-749-25
; Sequence 25, Application US/08710749
; Patent No. 5955089
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Becker, Robert
; TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,749
; FILING DATE: 20-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2074
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: n/a
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-710-749-23

Query Match
Best Local Similarity 100.0%; DB 2; Length 108;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 EDYKE 99
Db 68 EDYKE 73

RESULT 21
US-08-710-749-24
; Sequence 24, Application US/08710749
; Patent No. 5955089
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Becker, Robert
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 213:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-08528-213

Query Match 3.1%; Score 6; DB 5; Length 38;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 FIVGDW 67  
Db 19 FIVGDW 24

RESULT 17  
US-08-710-749-6

Sequence 6, Application US/08710749  
Patent No. 5955089

GENERAL INFORMATION:

APPLICANT: Briles, David E.

APPLICANT: Hollingshead, Susan

APPLICANT: Becker, Robert

TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE

TITLE OF INVENTION: PROTEINS

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Curtis, Morris & Safford

STREET: 530 Fifth Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/710,749

FILING DATE: 20-SEP-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454312-2074

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 101 amino acids

TYPE: amino acid

STRANDEDNESS: n/a

TOPOLOGY: linear

MOLECULE TYPE: amino acid

US-08-710-749-6

Query Match 3.1%; Score 6; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 EDYIKE 99  
Db 10 EDYIKE 15

RESULT 18

US-08-407-165-3

Sequence 3, Application US/08407165

Patent No. 6054280

GENERAL INFORMATION:

APPLICANT: LENNON, MARK A.

APPLICANT: FEIGUSON, KATHRYN M.

APPLICANT: SICHER, PAUL B.

APPLICANT: SCHLESSINGER, JOSEPH

TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT OF

TITLE OF INVENTION: PH DOMAIN SIGNAL TRANSDUCTION DISORDERS

FILE REFERENCE: 211/156

CURRENT APPLICATION NUMBER: US/08/407,165

CURRENT FILING DATE: 1995-03-20

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3

LENGTH: 105

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: Spectrin

US-08-407-165-3

Query Match 3.1%; Score 6; DB 3; Length 105;

Best Local Similarity 100.0%; Pred. No. 74;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 GFYKDA 81

Db 36 GFYKDA 41

RESULT 19

US-08-710-749-22

Sequence 22, Application US/08710749

Patent No. 5955089

GENERAL INFORMATION:

APPLICANT: Briles, David E.

APPLICANT: Hollingshead, Susan

APPLICANT: Becker, Robert

TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE

TITLE OF INVENTION: PROTEINS

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Curtis, Morris & Safford

STREET: 530 Fifth Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/710,749

FILING DATE: 20-SEP-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454312-2074

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 108 amino acids

TYPE: amino acid

STRANDEDNESS: n/a



```

; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 213:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-118-270-213
;
; Query Match 3.1%; Score 6; DB 1; Len
; Best Local Similarity 100.0%; Pred.No.29;
; Matches 6; Conservative 0; Mismatches 0;
;
Qy 62 FIVGDW 67
   |||||
Db 19 FIVGDW 24
;
;
RESULT 16
PCT-US93-08528-213
; Sequence 213, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PR
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2 PCT

```

```

RESULT 13
5240706-21
; Patent No. 5240706
; APPLICANT: FAULTS, DARYL
; TITLE OF INVENTION: INTRANASSAL ADMINISTRATION OF MYCOPLASMA
; HYOPNEUMONIAE ANTIGEN
; NUMBER OF SEQUENCES: 22
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/334,586
; FILING DATE: 07-APR-1989
; SEQ ID NO: 21:
; LENGTH: 24
5240706-21

Query Match          3.1%; Score 6; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 109 VLENPN 114
Db 18 VLENPN 23
|||||

RESULT 14
US-09-177-249-217
; Sequence 217, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramin
; APPLICANT: Margossian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
; FILE REFERENCE: 023070-086120US
; CURRENT APPLICATION NUMBER: US/09/177,249
; CURRENT FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: US 09/071,838
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 217
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-177-249-217

Query Match          3.1%; Score 6; DB 4; Length 25;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 KNSRIS 91
Db 9 KNSRIS 14
|||||

RESULT 15
US-08-118-270-213
; Sequence 213, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; APPLICANT: Schuster, David I.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:

```

Patent No. 5858701  
GENERAL INFORMATION:  
APPLICANT: WHITE, Morris F.  
APPLICANT: SUN, Xiao Jian  
APPLICANT: PIERCE, Jacalyn H.  
TITLE OF INVENTION: THE IRS FAMILY OF GENES  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,310A  
FILING DATE: 03-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Louis Myers  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: JDP-022  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1321 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-310A-16

Query Match 3.6%; Score 7; DB 2;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138  
|||||||  
Db 300 RSKSQSS 306

RESULT 11  
US-08-317-310A-64  
Sequence 64, Application US/08317310A  
Patent No. 5858701  
GENERAL INFORMATION:  
APPLICANT: WHITE, Morris F.  
APPLICANT: SUN, Xiao Jian  
APPLICANT: PIERCE, Jacalyn H.  
TITLE OF INVENTION: THE IRS FAMILY OF GENES  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,310A  
FILING DATE: 03-OCT-1994  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Louis Myers  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: JDP-022  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1321 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-310A-64

Query Match 3.6%; Score 7; DB 2;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138  
|||||||  
Db 300 RSKSQSS 306

RESULT 12  
PCT-US95-13041-16  
Sequence 16, Application PC/TUS9513041  
GENERAL INFORMATION:  
APPLICANT: WHITE, Morris F.  
APPLICANT: SUN, Xiao Jian  
APPLICANT: PIERCE, Jacalyn H.  
TITLE OF INVENTION: THE IRS FAMILY OF GENES  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 30 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13041  
FILING DATE: Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/317,310  
FILING DATE: 03-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Louis Myers  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: JDP-022PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1321 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-13041-16

Query Match 3.6%; Score 7; DB 2;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138  
|||||||  
Db 300 RSKSQSS 306

Query Match 3.6%; Score 7; DB 5;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138  
|||||||  
Db 300 RSKSQSS 306

ADDRESSEE: NO. 5827730o No. 5827730disk of No. 5827730th America  
STREET: 405 Lexington Avenue  
CITY: New York

US-09-085-761A-2  
; Sequence 2, Application US/09085761A  
; Patent No. 6335178  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Joel H.  
; APPLICANT: Turner, Raymond J.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN  
; TITLE OF INVENTION: SECRETION  
; NUMBER OF SEQUENCES: 77  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085,761A  
; FILING DATE: 28-MAY-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: UALB-03356  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 284 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-09-085-761A-2

Query Match 3.6%; Score 7; DB 4; Length 284;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 SKSOSSK 139  
Db 275 SKSOSSK 281

RESULT 5  
US-08-937-067-6  
; Sequence 6, Application US/08937067  
; Patent No. 6433155  
; GENERAL INFORMATION:  
; APPLICANT: Umansky, Samuil  
; APPLICANT: Melkonyan, Hovsep  
; TITLE OF INVENTION: A FAMILY OF GENES ENCODING  
; TITLE OF INVENTION: APOPTOSIS-RELATED PEPTIDES; PEPTIDES ENCODED THEREBY AND  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/937,067  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lehnhardt, Susan K.  
; REGISTRATION NUMBER: 33,943  
; REFERENCE/DOCKET NUMBER: 23647-20018.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 813-5600  
; TELEFAX: (650) 494-0792  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 317 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-937-067-6

Query Match 3.6%; Score 7; DB 4; Length 317;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 159 KRKDTKF 165  
Db 236 KRKDTKF 242

RESULT 6  
US-08-317-310A-15  
; Sequence 15, Application US/08317310A  
; Patent No. 5858701  
; GENERAL INFORMATION:  
; APPLICANT: WHITE, Morris F.  
; APPLICANT: SUN, Xiao Jian  
; APPLICANT: PIERCE, Jacalyn H.  
; TITLE OF INVENTION: THE IRS FAMILY OF GENES  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,310A  
; FILING DATE: 03-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Louis Myers  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: JDP-022  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1234 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
US-08-317-310A-15

Query Match 3.6%; Score 7; DB 2; Length 1234;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

977 5 2.6 335 4 US-09-232-191-21 Sequence 21, Appl  
978 5 2.6 335 4 US-09-232-200-21 Sequence 21, Appl  
979 5 2.6 335 4 US-09-232-197-21 Sequence 21, Appl  
980 5 2.6 335 4 US-09-180-100-20 Sequence 20, Appl  
981 5 2.6 335 4 US-09-232-201-21 Sequence 21, Appl  
982 5 2.6 335 4 US-09-565-918-3 Sequence 3, Appl  
983 5 2.6 335 5 PCT-US95-17083-2 Sequence 7, Appl  
984 5 2.6 336 1 US-07-928-462-2 Sequence 2, Appl  
985 5 2.6 336 4 US-08-273-247-2 Sequence 2, Appl  
986 5 2.6 337 1 US-07-966-187-25 Sequence 25, Appl  
987 5 2.6 337 1 US-08-466-033-162 Sequence 162, App  
988 5 2.6 337 2 US-08-444-733-162 Sequence 162, App  
989 5 2.6 337 2 US-08-464-134-162 Sequence 162, App  
990 5 2.6 337 2 US-08-461-361-162 Sequence 162, App  
991 5 2.6 337 2 US-08-485-910-162 Sequence 162, App  
992 5 2.6 337 3 US-09-032-372-2 Sequence 2, Appl  
993 5 2.6 337 3 US-08-686-528A-2 Sequence 2, Appl  
994 5 2.6 337 3 US-09-190-965-1 Sequence 1, Appl  
995 5 2.6 337 4 US-09-456-287-2 Sequence 2, Appl  
996 5 2.6 337 4 US-09-470-253-1 Sequence 1, Appl  
997 5 2.6 338 1 US-07-966-187-3 Sequence 3, Appl  
998 5 2.6 338 1 US-07-966-187-16 Sequence 16, Appl  
999 5 2.6 338 4 US-08-448-398-13 Sequence 13, Appl  
1000

## ALIGNMENTS

RESULT 1  
US-09-134-001C-4894  
; Sequence 4894, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4894  
; LENGTH: 132  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4894

Query Match 3.6%; Score 7; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105  
Db 78 EYCNFGC 84

RESULT 2  
US-09-297-937C-7  
; Sequence 7, Application US/09297937C  
; Patent No. 6337199  
; GENERAL INFORMATION:  
; APPLICANT: YUN, Do Young  
; APPLICANT: PAN, Jae Gu  
; TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence  
; TITLE OF INVENTION: Encoding the Same and Production of 2-Keto-D-Gluconate  
; TITLE OF INVENTION: Using Transformed Recombinant E. Coli  
; FILE REFERENCE: P66159US0  
; CURRENT APPLICATION NUMBER: US/09/297,937C  
; CURRENT FILING DATE: 1999-05-11

; PRIOR APPLICATION NUMBER: PCT/KR98/00296  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: KR 97-48802  
; PRIOR FILING DATE: 1997-09-25  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Erwinia cyripedii  
US-09-297-937C-7

Query Match 3.6%; Score 7; DB 4; Length 220;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 QSSKSQT 142  
Db 45 QSSKSQT 51

RESULT 3  
US-09-053-197A-2  
; Sequence 2, Application US/09053197A  
; Patent No. 6022952  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Joel H.  
; APPLICANT: Turner, Raymond J.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN  
; TITLE OF INVENTION: SECRETION  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/053,197A  
; FILING DATE: 01-APR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MacKnight, Kamrin T.  
; REGISTRATION NUMBER: 38,230  
; REFERENCE/DOCKET NUMBER: UALB-03293  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 284 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-09-053-197A-2

Query Match 3.6%; Score 7; DB 3; Length 284;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 SKSQSSK 139  
Db 275 SKSQSSK 281

RESULT 4

831	5	2.6	281	4	US-09-479-524-3	Sequence 3, Appli	904	5	2.6	310	3	US-09-063-869-4	Sequence 4, Appli
832	5	2.6	281	4	US-08-339-214-8	Sequence 8, Appli	905	5	2.6	313	3	US-08-686-528A-3	Sequence 3, Appli
833	5	2.6	281	4	US-08-339-214-30	Sequence 30, Appli	906	5	2.6	313	3	US-09-456-287-3	Sequence 3, Appli
834	5	2.6	281	4	US-09-589-287B-6	Sequence 6, Appli	907	5	2.6	314	1	US-08-444-231-19	Sequence 19, Appli
835	5	2.6	281	4	US-09-157-864-10	Sequence 10, Appli	908	5	2.6	314	1	US-08-152-443A-19	Sequence 19, Appli
836	5	2.6	281	4	US-09-006-755B-1	Sequence 1, Appli	909	5	2.6	314	5	PCT-US95-17083-4	Sequence 4, Appli
837	5	2.6	281	5	PCT-US95-00362-2	Sequence 2, Appli	910	5	2.6	315	3	US-09-135-639-4	Sequence 4, Appli
838	5	2.6	282	1	US-07-712-476A-5	Sequence 5, Appli	911	5	2.6	315	4	US-09-134-001C-4380	Sequence 4380, Ap
839	5	2.6	283	1	US-08-583-672-2	Sequence 2, Appli	912	5	2.6	316	2	US-08-846-762-4	Sequence 4, Appli
840	5	2.6	283	2	US-08-202-044-2	Sequence 2, Appli	913	5	2.6	317	2	US-08-973-275-1	Sequence 1, Appli
841	5	2.6	283	4	US-08-751-344B-2	Sequence 2, Appli	914	5	2.6	318	3	US-08-859-167-4	Sequence 4, Appli
842	5	2.6	284	2	US-08-320-148B-2	Sequence 2, Appli	915	5	2.6	318	3	US-08-859-167-6	Sequence 6, Appli
843	5	2.6	284	3	US-08-589-028-6	Sequence 6, Appli	916	5	2.6	318	3	US-09-109-273-4	Sequence 4, Appli
844	5	2.6	284	3	US-08-491-954-3	Sequence 3, Appli	917	5	2.6	318	3	US-09-109-273-6	Sequence 6, Appli
845	5	2.6	284	3	US-08-784-582-6	Sequence 6, Appli	918	5	2.6	318	4	US-08-949-246-2	Sequence 2, Appli
846	5	2.6	284	4	US-08-785-271-6	Sequence 6, Appli	919	5	2.6	318	4	US-09-276-993-4	Sequence 4, Appli
847	5	2.6	284	4	US-09-031-898-2	Sequence 2, Appli	920	5	2.6	318	4	US-09-276-993-6	Sequence 6, Appli
848	5	2.6	285	2	US-08-418-071-17	Sequence 17, Appli	921	5	2.6	318	4	US-09-134-001C-3603	Sequence 3603, Ap
849	5	2.6	286	5	PCT-US92-00282-9	Sequence 9, Appli	922	5	2.6	320	1	US-08-565-386-17	Sequence 17, Appli
850	5	2.6	287	3	US-08-815-190A-16	Sequence 16, Appli	923	5	2.6	321	4	US-08-506-296B-54	Sequence 54, Appli
851	5	2.6	287	3	US-08-549-515-10	Sequence 10, Appli	924	5	2.6	321	4	US-09-134-001C-2890	Sequence 2890, Ap
852	5	2.6	287	4	US-08-862-124-17	Sequence 17, Appli	925	5	2.6	322	1	US-08-097-828-1	Sequence 1, Appli
853	5	2.6	289	5	PCT-US92-00282-13	Sequence 13, Appli	926	5	2.6	322	1	US-08-036-210-11	Sequence 11, Appli
854	5	2.6	289	5	PCT-US92-00282-15	Sequence 15, Appli	927	5	2.6	322	1	US-08-480-756-1	Sequence 1, Appli
855	5	2.6	290	2	US-08-903-801-1	Sequence 1, Appli	928	5	2.6	322	2	US-08-462-403-1	Sequence 1, Appli
856	5	2.6	290	4	US-09-295-055-1	Sequence 1, Appli	929	5	2.6	322	2	US-08-286-819A-2	Sequence 2, Appli
857	5	2.6	292	2	US-08-701-191A-40	Sequence 40, Appli	930	5	2.6	322	2	US-08-449-609-11	Sequence 11, Appli
858	5	2.6	294	2	US-08-424-641B-6	Sequence 6, Appli	931	5	2.6	322	3	US-08-980-357-2	Sequence 2, Appli
859	5	2.6	294	2	US-08-820-980-6	Sequence 6, Appli	932	5	2.6	322	3	PCT-US92-00331-1	Sequence 1, Appli
860	5	2.6	294	2	US-08-826-439-6	Sequence 6, Appli	933	5	2.6	322	5	PCT-US93-10419-1	Sequence 1, Appli
861	5	2.6	294	4	US-09-131-028A-9	Sequence 9, Appli	934	5	2.6	326	2	US-08-986-217-3	Sequence 2, Appli
862	5	2.6	294	4	US-09-131-028A-11	Sequence 11, Appli	935	5	2.6	326	4	US-09-066-046-29	Sequence 29, Appli
863	5	2.6	295	1	US-08-149-091-1	Sequence 1, Appli	936	5	2.6	326	4	US-09-066-047-15	Sequence 15, Appli
864	5	2.6	295	2	US-08-637-761-1	Sequence 1, Appli	937	5	2.6	327	1	US-08-080-386-2	Sequence 2, Appli
865	5	2.6	295	4	US-09-134-001C-5643	Sequence 5643, Ap	938	5	2.6	327	2	US-08-390-000A-2	Sequence 2, Appli
866	5	2.6	295	5	PCT-US94-12672-1	Sequence 1, Appli	939	5	2.6	327	4	US-09-446-504-3	Sequence 3, Appli
867	5	2.6	296	1	US-07-712-476A-1	Sequence 1, Appli	940	5	2.6	327	4	US-09-712-266-3	Sequence 3, Appli
868	5	2.6	296	4	US-08-944-604-20	Sequence 20, Appli	941	5	2.6	327	4	US-09-583-492-10	Sequence 10, Appli
869	5	2.6	297	2	US-08-874-347-23	Sequence 23, Appli	942	5	2.6	328	1	US-08-423-691-2	Sequence 2, Appli
870	5	2.6	297	2	US-09-006-535-4	Sequence 4, Appli	943	5	2.6	328	1	US-08-080-386-4	Sequence 4, Appli
871	5	2.6	297	3	US-09-093-522-23	Sequence 23, Appli	944	5	2.6	328	2	US-08-390-000A-4	Sequence 4, Appli
872	5	2.6	298	1	US-08-149-091-6	Sequence 6, Appli	945	5	2.6	328	3	US-08-867-260-2	Sequence 2, Appli
873	5	2.6	298	2	US-08-637-761-6	Sequence 6, Appli	946	5	2.6	328	4	US-09-501-192-9	Sequence 9, Appli
874	5	2.6	298	2	US-09-006-535-3	Sequence 3, Appli	947	5	2.6	328	4	US-09-428-747-4	Sequence 4, Appli
875	5	2.6	298	5	PCT-US94-12672-6	Sequence 6, Appli	948	5	2.6	328	4	US-09-615-192A-274	Sequence 274, App
876	5	2.6	299	4	US-09-134-001C-5227	Sequence 5227, Ap	949	5	2.6	329	1	US-07-991-587A-7	Sequence 7, Appli
877	5	2.6	300	2	US-08-828-010-2	Sequence 2, Appli	950	5	2.6	329	1	US-08-309-985-7	Sequence 7, Appli
878	5	2.6	300	6	5340934-6	Patent No. 5340934	951	5	2.6	329	4	US-09-189-527-4	Sequence 4, Appli
879	5	2.6	301	4	US-08-944-604-18	Sequence 18, Appli	952	5	2.6	330	2	US-08-525-864A-4	Sequence 4, Appli
880	5	2.6	301	4	US-09-288-143-98	Sequence 98, Appli	953	5	2.6	331	1	US-08-466-033-166	Sequence 166, App
881	5	2.6	302	4	US-09-457-046B-18	Sequence 18, Appli	954	5	2.6	331	2	US-08-444-733-166	Sequence 166, App
882	5	2.6	303	1	US-08-109-391A-2	Sequence 2, Appli	955	5	2.6	331	2	US-08-464-134-166	Sequence 166, App
883	5	2.6	303	1	US-08-459-019A-2	Sequence 2, Appli	956	5	2.6	331	2	US-08-461-361-166	Sequence 166, App
884	5	2.6	303	2	US-08-460-428A-2	Sequence 2, Appli	957	5	2.6	331	2	US-08-485-910-166	Sequence 166, App
885	5	2.6	303	3	US-08-458-860A-2	Sequence 2, Appli	958	5	2.6	331	2	US-09-986-217-7	Sequence 7, Appli
886	5	2.6	303	4	US-09-105-390-38	Sequence 38, Appli	959	5	2.6	331	4	US-09-086-483A-3	Sequence 3, Appli
887	5	2.6	304	1	US-07-851-976B-8	Sequence 8, Appli	960	5	2.6	331	4	US-09-199-637A-333	Sequence 333, App
888	5	2.6	304	1	US-08-291-609-8	Sequence 8, Appli	961	5	2.6	332	2	US-08-446-875-12	Sequence 12, Appli
889	5	2.6	304	1	US-08-401-136-8	Sequence 8, Appli	962	5	2.6	332	2	US-08-102-385G-12	Sequence 12, Appli
890	5	2.6	304	3	US-08-850-554-8	Sequence 8, Appli	963	5	2.6	332	4	US-09-105-390-54	Sequence 54, Appli
891	5	2.6	304	4	US-08-862-124-14	Sequence 14, Appli	964	5	2.6	333	4	US-08-961-083-54	Sequence 54, Appli
892	5	2.6	304	4	US-09-739-455-13	Sequence 13, Appli	965	5	2.6	334	1	US-08-118-270-73	Sequence 73, Appli
893	5	2.6	304	4	US-09-739-455-23	Sequence 23, Appli	966	5	2.6	334	1	US-08-036-210-7	Sequence 7, Appli
894	5	2.6	305	4	US-09-134-001C-3192	Sequence 3192, Ap	967	5	2.6	334	2	US-08-449-609-7	Sequence 7, Appli
895	5	2.6	305	4	US-09-451-900-1	Sequence 1, Appli	968	5	2.6	334	4	US-09-218-363-11	Sequence 11, Appli
896	5	2.6	306	4	US-09-921-318-1	Sequence 1, Appli	969	5	2.6	334	5	PCT-US93-08528-73	Sequence 73, Appli
897	5	2.6	307	4	US-08-949-246-4	Sequence 4, Appli	970	5	2.6	334	6	5290690-11	Patent No. 5290590
898	5	2.6	307	4	US-09-134-001C-5144	Sequence 5144, Ap	971	5	2.6	335	2	US-08-219-237B-2	Sequence 2, Appli
899	5	2.6	308	4	US-08-952-089A-13	Sequence 13, Appli	972	5	2.6	335	2	US-08-409-338-1	Sequence 1, Appli
900	5	2.6	309	1	US-07-966-187-1	Sequence 1, Appli	973	5	2.6	335	4	US-08-815-469-6	Sequence 6, Appli
901	5	2.6	309	3	US-08-605-284B-11	Sequence 11, Appli	974	5	2.6	335	4	US-09-230-640-2	Sequence 2, Appli
902	5	2.6	309	4	US-08-448-398-11	Sequence 11, Appli	975	5	2.6	335	4	US-09-006-353A-7	Sequence 7, Appli
903	5	2.6	310	2	US-08-695-355-4	Sequence 4, Appli	976	5	2.6	335	4	US-08-468-560C-2	Sequence 2, Appli

685	5	2.6	192	4	US-09-303-120B-8	Sequence 8, Appl	758	5	2.6	243	4	US-09-181-958-1	Sequence 1, Appl
686	5	2.6	192	4	US-09-820-576-8	Sequence 8, Appl	759	5	2.6	243	4	US-09-370-838-39	Sequence 39, Appl
687	5	2.6	192	4	US-09-218-363-18	Sequence 18, Appl	760	5	2.6	243	4	US-09-370-838-42	Sequence 42, Appl
688	5	2.6	198	3	US-08-737-248-6	Sequence 6, Appl	761	5	2.6	244	4	US-09-370-838-43	Sequence 43, Appl
689	5	2.6	199	2	US-08-405-175A-3	Sequence 3, Appl	762	5	2.6	244	4	US-09-370-838-46	Sequence 46, Appl
690	5	2.6	199	4	US-09-497-779A-6	Sequence 6, Appl	763	5	2.6	245	4	US-09-370-838-40	Sequence 40, Appl
691	5	2.6	200	2	US-08-405-175A-4	Sequence 4, Appl	764	5	2.6	246	4	US-08-822-774-45	Sequence 45, Appl
692	5	2.6	200	4	US-08-858-207A-390	Sequence 390, App	765	5	2.6	246	4	US-09-632-711-45	Sequence 45, Appl
693	5	2.6	201	2	US-09-933-750C-21	Sequence 21, Appl	766	5	2.6	246	4	US-09-632-703B-45	Sequence 45, Appl
694	5	2.6	201	4	US-09-234-613-21	Sequence 21, Appl	767	5	2.6	246	4	US-09-632-702-45	Sequence 45, Appl
695	5	2.6	202	4	US-09-370-838-38	Sequence 38, Appl	768	5	2.6	248	4	US-09-163-285-2	Sequence 2, Appl
696	5	2.6	203	4	US-09-134-001C-3111	Sequence 3111, Ap	769	5	2.6	248	4	US-09-134-001C-3731	Sequence 3731, Ap
697	5	2.6	204	2	US-08-516-801-3	Sequence 3, Appl	770	5	2.6	248	4	US-09-452-239-26	Sequence 26, Appl
698	5	2.6	204	4	US-08-248-355-3	Sequence 3, Appl	771	5	2.6	249	1	US-08-680-726A-88	Sequence 88, Appl
699	5	2.6	204	5	PCT-US95-06683-3	Sequence 3, Appl	772	5	2.6	249	4	US-09-092-409-88	Sequence 88, Appl
700	5	2.6	205	2	US-08-684-024-1	Sequence 1, Appl	773	5	2.6	249	4	US-09-578-303-2	Sequence 2, Appl
701	5	2.6	205	2	US-08-684-024-6	Sequence 6, Appl	774	5	2.6	251	2	US-08-401-068-12	Sequence 12, Appl
702	5	2.6	205	2	US-08-684-024-7	Sequence 7, Appl	775	5	2.6	251	2	US-08-846-338-12	Sequence 12, Appl
703	5	2.6	205	3	US-09-145-868-1	Sequence 1, Appl	776	5	2.6	251	3	US-08-411-768B-3	Sequence 3, Appl
704	5	2.6	205	3	US-09-145-868-6	Sequence 6, Appl	777	5	2.6	255	4	US-09-813-817-2	Sequence 2, Appl
705	5	2.6	205	3	US-09-145-868-7	Sequence 7, Appl	778	5	2.6	255	4	US-09-978-197-2	Sequence 2, Appl
706	5	2.6	208	4	US-09-134-001C-5600	Sequence 5600, Ap	779	5	2.6	258	1	US-08-149-091-2	Sequence 2, Appl
707	5	2.6	208	4	US-09-549-831-2	Sequence 2, Appl	780	5	2.6	258	2	US-08-637-761-2	Sequence 2, Appl
708	5	2.6	208	4	US-09-164-193-8	Sequence 8, Appl	781	5	2.6	258	4	US-09-497-779A-9	Sequence 9, Appl
709	5	2.6	209	4	US-09-221-448A-8	Sequence 8, Appl	782	5	2.6	258	4	US-09-134-001C-4806	Sequence 4806, Ap
710	5	2.6	209	6	5427917-2	Patent No. 5427917	783	5	2.6	258	4	US-09-134-001C-5593	Sequence 5593, Ap
711	5	2.6	210	2	US-08-781-560-4	Sequence 4, Appl	784	5	2.6	258	5	PCT-US94-12672-2	Sequence 2, Appl
712	5	2.6	210	2	US-08-933-750C-6	Sequence 6, Appl	785	5	2.6	259	2	US-08-402-804-8	Sequence 8, Appl
713	5	2.6	210	2	US-08-747-536-2	Sequence 2, Appl	786	5	2.6	260	2	US-08-763-919-2	Sequence 2, Appl
714	5	2.6	210	2	US-08-747-536-4	Sequence 4, Appl	787	5	2.6	260	4	US-08-560-729-2	Sequence 2, Appl
715	5	2.6	210	2	US-08-747-536-6	Sequence 6, Appl	788	5	2.6	261	4	US-09-721-259-2	Sequence 2, Appl
716	5	2.6	210	2	US-08-747-536-8	Sequence 8, Appl	789	5	2.6	261	4	US-08-989-478-16	Sequence 16, Appl
717	5	2.6	210	3	US-08-792-014-4	Sequence 4, Appl	790	5	2.6	261	3	US-08-996-685-16	Sequence 16, Appl
718	5	2.6	210	4	US-09-234-613-6	Sequence 6, Appl	791	5	2.6	261	4	US-09-097-889-19	Sequence 19, Appl
719	5	2.6	210	4	US-08-297-431B-34	Sequence 34, Appl	792	5	2.6	262	4	US-08-491-944-2	Sequence 2, Appl
720	5	2.6	210	4	US-09-443-948-4	Sequence 4, Appl	793	5	2.6	263	2	US-08-149-097D-28	Sequence 28, Appl
721	5	2.6	214	2	US-08-480-753-6	Sequence 6, Appl	794	5	2.6	263	4	US-08-818-112-91	Sequence 91, Appl
722	5	2.6	214	2	US-08-846-790A-1	Sequence 1, Appl	795	5	2.6	263	4	US-08-818-111-92	Sequence 92, Appl
723	5	2.6	214	3	US-08-935-333-1	Sequence 1, Appl	796	5	2.6	263	4	US-09-056-556-91	Sequence 91, Appl
724	5	2.6	214	3	US-09-041-889-11	Sequence 11, Appl	797	5	2.6	263	4	US-09-072-596-92	Sequence 92, Appl
725	5	2.6	214	3	US-08-837-058-11	Sequence 11, Appl	798	5	2.6	264	2	US-08-728-521-1	Sequence 1, Appl
726	5	2.6	215	2	US-08-531-525-49	Sequence 49, Appl	799	5	2.6	264	4	US-09-212-146-1	Sequence 1, Appl
727	5	2.6	215	2	US-08-718-270A-49	Sequence 49, Appl	800	5	2.6	264	4	US-09-134-001C-3780	Sequence 3780, Ap
728	5	2.6	215	4	US-09-240-410-4	Sequence 4, Appl	801	5	2.6	265	4	US-09-813-817-4	Sequence 4, Appl
729	5	2.6	218	1	US-07-607-538C-4	Sequence 4, Appl	802	5	2.6	265	4	US-09-134-001C-3336	Sequence 3336, Ap
730	5	2.6	218	2	US-08-162-402B-4	Sequence 4, Appl	803	5	2.6	265	4	US-09-978-197-4	Sequence 4, Appl
731	5	2.6	218	4	US-08-914-375C-6	Sequence 6, Appl	804	5	2.6	269	4	US-09-082-649B-82	Sequence 82, Appl
732	5	2.6	219	4	US-09-134-001C-5651	Sequence 5651, Ap	805	5	2.6	271	1	US-07-977-630-5	Sequence 5, Appl
733	5	2.6	221	2	US-08-861-269-6	Sequence 6, Appl	806	5	2.6	271	2	US-08-850-880-6	Sequence 6, Appl
734	5	2.6	221	2	US-09-134-596-6	Sequence 6, Appl	807	5	2.6	271	2	US-08-944-916-6	Sequence 6, Appl
735	5	2.6	221	3	US-09-293-273-6	Sequence 6, Appl	808	5	2.6	271	2	US-08-814-877-6	Sequence 6, Appl
736	5	2.6	223	4	US-09-230-196-17	Sequence 17, Appl	809	5	2.6	271	4	US-08-760-615-8	Sequence 8, Appl
737	5	2.6	223	4	US-09-230-196-24	Sequence 24, Appl	810	5	2.6	271	4	US-09-131-028A-5	Sequence 5, Appl
738	5	2.6	226	3	US-08-966-318-1	Sequence 1, Appl	811	5	2.6	271	4	US-09-131-028A-7	Sequence 7, Appl
739	5	2.6	226	4	US-09-216-619-1	Sequence 1, Appl	812	5	2.6	271	4	US-09-272-432A-6	Sequence 6, Appl
740	5	2.6	226	4	US-09-163-285-4	Sequence 4, Appl	813	5	2.6	272	1	US-08-252-995D-14	Sequence 14, Appl
741	5	2.6	226	4	US-09-134-001C-5520	Sequence 5520, Ap	814	5	2.6	272	2	US-08-834-108-14	Sequence 14, Appl
742	5	2.6	227	4	US-09-134-001C-4836	Sequence 4836, Ap	815	5	2.6	273	2	US-08-989-386-8	Sequence 8, Appl
743	5	2.6	228	2	US-08-408-095-21	Sequence 21, Appl	816	5	2.6	273	2	US-09-134-001C-3917	Sequence 3917, Ap
744	5	2.6	229	2	US-08-726-306A-23	Sequence 23, Appl	817	5	2.6	275	2	US-08-645-193B-19	Sequence 19, Appl
745	5	2.6	229	3	US-08-840-146-20	Sequence 20, Appl	818	5	2.6	276	1	US-07-952-817-27	Sequence 27, Appl
746	5	2.6	234	3	US-09-360-220-20	Sequence 20, Appl	819	5	2.6	278	2	US-08-685-992-34	Sequence 34, Appl
747	5	2.6	234	1	US-08-684-862-3	Sequence 3, Appl	820	5	2.6	278	2	US-09-144-925-34	Sequence 34, Appl
748	5	2.6	234	2	US-08-738-413B-12	Sequence 12, Appl	821	5	2.6	278	4	US-08-339-214-16	Sequence 16, Appl
749	5	2.6	234	4	US-09-134-001C-3195	Sequence 3195, Ap	822	5	2.6	278	4	US-08-339-214-26	Sequence 26, Appl
750	5	2.6	235	3	US-08-812-586-16	Sequence 16, Appl	823	5	2.6	278	6	5340934-8	Patent No. 5340934
751	5	2.6	235	4	US-09-029-785-1	Sequence 1, Appl	824	5	2.6	279	4	US-09-330-970-5	Sequence 5, Appl
752	5	2.6	235	4	US-09-370-838-37	Sequence 37, Appl	825	5	2.6	279	4	US-08-339-214-24	Sequence 24, Appl
753	5	2.6	239	2	US-08-286-819A-44	Sequence 44, Appl	826	5	2.6	279	5	US-08-339-214-32	Sequence 32, Appl
754	5	2.6	239	3	US-08-980-357-44	Sequence 44, Appl	827	5	2.6	279	5	PCT-US95-00362-5	Sequence 5, Appl
755	5	2.6	239	3	US-08-812-586-29	Sequence 29, Appl	828	5	2.6	281	2	US-08-810-453-2	Sequence 2, Appl
756	5	2.6	240	4	US-09-194-905-8	Sequence 8, Appl	829	5	2.6	281	3	US-08-815-190A-2	Sequence 2, Appl
757	5	2.6	241	4	US-09-387-372-3	Sequence 3, Appl	830	5	2.6	281	4	US-09-290-640-25	Sequence 25, Appl

539	5	2.6	123	4	US-09-156-316-10	Sequence 10, Appl	612	5	2.6	155	2	US-08-484-101B-12	Sequence 12, Appl
540	5	2.6	124	1	US-09-134-001C-5344	Sequence 5344, Ap	613	5	2.6	155	4	US-08-714-524D-12	Sequence 12, Appl
541	5	2.6	125	1	US-08-197-496A-2	Sequence 2, Appli	614	5	2.6	156	2	US-09-070-060-7	Sequence 7, Appli
542	5	2.6	125	1	US-08-275-370-2	Sequence 2, Appli	615	5	2.6	156	3	US-09-051-969A-3	Sequence 3, Appli
543	5	2.6	125	1	US-08-462-949-24	Sequence 24, Appl	616	5	2.6	156	3	US-09-051-969A-4	Sequence 4, Appli
544	5	2.6	125	1	US-08-367-968-2	Sequence 2, Appli	617	5	2.6	156	3	US-09-357-746-7	Sequence 7, Appli
545	5	2.6	125	1	US-08-023-764B-24	Sequence 24, Appl	618	5	2.6	156	4	US-09-615-192A-340	Sequence 340, App
546	5	2.6	125	1	US-08-665-484-2	Sequence 2, Appli	619	5	2.6	156	4	US-09-053-197A-19	Sequence 19, Appl
547	5	2.6	125	3	US-08-665-259-35	Sequence 35, Appl	620	5	2.6	158	3	US-08-946-329A-66	Sequence 66, Appl
548	5	2.6	125	3	US-08-762-500-35	Sequence 35, Appl	621	5	2.6	158	4	US-09-091-725-25	Sequence 25, Appl
549	5	2.6	126	2	US-08-341-843B-36	Sequence 36, Appl	622	5	2.6	158	4	US-09-085-761A-19	Sequence 19, Appl
550	5	2.6	126	2	US-08-427-497E-41	Sequence 41, Appl	623	5	2.6	159	2	US-08-829-110-1	Sequence 1, Appli
551	5	2.6	126	4	US-09-209-525-45	Sequence 45, Appl	624	5	2.6	159	2	US-08-162-402B-16	Sequence 16, Appli
552	5	2.6	127	2	US-08-476-176B-6	Sequence 6, Appli	625	5	2.6	160	4	US-08-858-207A-367	Sequence 367, App
553	5	2.6	127	2	US-08-476-176B-8	Sequence 8, Appli	626	5	2.6	160	4	US-09-370-838-205	Sequence 205, App
554	5	2.6	127	2	US-08-476-176B-10	Sequence 10, Appl	627	5	2.6	161	2	US-08-360-606B-33	Sequence 33, Appl
555	5	2.6	127	3	US-08-127-721A-6	Sequence 6, Appli	628	5	2.6	164	4	US-09-134-001C-4630	Sequence 4630, Ap
556	5	2.6	127	3	US-08-127-721A-8	Sequence 8, Appli	629	5	2.6	165	2	US-08-718-599-1	Sequence 1, Appli
557	5	2.6	127	3	US-08-127-721A-10	Sequence 10, Appl	630	5	2.6	165	2	US-08-587-680A-27	Sequence 27, Appl
558	5	2.6	127	3	US-08-485-246A-6	Sequence 6, Appli	631	5	2.6	165	3	US-08-765-381-4	Sequence 4, Appli
559	5	2.6	127	3	US-08-485-246A-8	Sequence 8, Appli	632	5	2.6	166	3	US-08-765-381-12	Sequence 12, Appl
560	5	2.6	127	3	US-08-485-246A-10	Sequence 10, Appl	633	5	2.6	166	4	US-09-134-001C-4279	Sequence 4279, Ap
561	5	2.6	128	2	US-08-557-309B-41	Sequence 41, Appl	634	5	2.6	168	1	US-08-145-995A-12	Sequence 12, Appl
562	5	2.6	128	3	US-08-834-306-41	Sequence 41, Appl	635	5	2.6	168	2	US-08-451-747-12	Sequence 12, Appl
563	5	2.6	128	4	US-08-993-674A-41	Sequence 41, Appl	636	5	2.6	168	3	US-08-946-329A-67	Sequence 67, Appl
564	5	2.6	128	4	US-09-256-976-41	Sequence 41, Appl	637	5	2.6	168	3	US-09-134-852-12	Sequence 12, Appl
565	5	2.6	130	4	US-08-339-214-28	Sequence 28, Appl	638	5	2.6	174	4	US-09-149-476-424	Sequence 424, App
566	5	2.6	134	1	US-08-405-034-4	Sequence 4, Appli	639	5	2.6	176	1	US-08-036-210-9	Sequence 9, Appli
567	5	2.6	136	4	US-08-339-214-83	Sequence 83, Appl	640	5	2.6	176	2	US-08-449-609-9	Sequence 9, Appli
568	5	2.6	137	3	US-08-685-871-60	Sequence 60, Appl	641	5	2.6	178	4	US-08-339-214-84	Sequence 84, Appl
569	5	2.6	137	4	US-08-339-214-20	Sequence 20, Appl	642	5	2.6	178	4	US-09-134-001C-3412	Sequence 3412, Ap
570	5	2.6	137	4	US-08-339-214-10	Sequence 10, Appl	643	5	2.6	179	3	US-08-649-100-9	Sequence 9, Appli
571	5	2.6	137	4	US-08-339-214-18	Sequence 18, Appl	644	5	2.6	179	4	US-08-874-102-41	Sequence 41, Appl
572	5	2.6	137	4	US-08-339-214-82	Sequence 82, Appl	645	5	2.6	179	4	US-08-874-102-44	Sequence 44, Appl
573	5	2.6	138	3	US-08-930-894-5	Sequence 5, Appli	646	5	2.6	179	4	US-08-339-214-6	Sequence 6, Appli
574	5	2.6	138	4	US-08-339-214-4	Sequence 4, Appli	647	5	2.6	179	4	US-08-339-214-14	Sequence 14, Appl
575	5	2.6	138	4	US-08-339-214-12	Sequence 12, Appl	648	5	2.6	179	4	US-08-339-214-22	Sequence 22, Appl
576	5	2.6	138	4	US-08-339-214-20	Sequence 20, Appl	649	5	2.6	179	4	US-08-984-919A-41	Sequence 41, Appl
577	5	2.6	138	4	US-08-339-214-81	Sequence 81, Appl	650	5	2.6	179	4	US-08-984-919A-44	Sequence 44, Appl
578	5	2.6	139	3	US-08-753-007A-33	Sequence 33, Appl	651	5	2.6	179	5	PCT-US93-04365-7	Sequence 7, Appli
579	5	2.6	139	4	US-09-398-496-33	Sequence 33, Appl	652	5	2.6	180	4	US-09-134-001C-2966	Sequence 2966, Ap
580	5	2.6	139	4	US-08-339-214-80	Sequence 80, Appl	653	5	2.6	181	3	US-08-753-007A-4	Sequence 4, Appli
581	5	2.6	139	4	US-08-936-165A-409	Sequence 409, App	654	5	2.6	181	4	US-09-398-496-4	Sequence 4, Appli
582	5	2.6	139	4	US-09-134-001C-5124	Sequence 5124, Ap	655	5	2.6	182	1	US-07-991-867B-10	Sequence 10, Appl
583	5	2.6	140	2	US-08-477-451-39	Sequence 39, Appl	656	5	2.6	182	1	US-08-117-083-10	Sequence 10, Appl
584	5	2.6	141	2	US-08-658-639-14	Sequence 14, Appl	657	5	2.6	182	1	US-08-107-755A-10	Sequence 10, Appl
585	5	2.6	141	4	US-08-944-604-14	Sequence 14, Appl	658	5	2.6	182	2	US-08-544-332-10	Sequence 10, Appl
586	5	2.6	141	4	US-09-286-529-22	Sequence 22, Appl	659	5	2.6	182	4	US-09-134-001C-3742	Sequence 3742, Ap
587	5	2.6	141	4	US-08-339-214-79	Sequence 79, Appl	660	5	2.6	182	4	US-09-370-861A-10	Sequence 10, Appl
588	5	2.6	142	4	US-09-134-001C-3473	Sequence 3473, Ap	661	5	2.6	183	3	US-09-122-443-11	Sequence 11, Appl
589	5	2.6	143	1	US-07-661-610C-8	Sequence 8, Appli	662	5	2.6	186	1	US-08-928-443-3	Sequence 3, Appli
590	5	2.6	143	4	US-09-325-932A-197	Sequence 197, App	663	5	2.6	186	3	US-09-129-055-3	Sequence 3, Appli
591	5	2.6	145	2	US-08-788-584-5	Sequence 5, Appli	664	5	2.6	187	2	US-08-177-109A-61	Sequence 61, Appl
592	5	2.6	145	3	US-08-946-914-12	Sequence 12, Appl	665	5	2.6	187	2	US-08-687-706-61	Sequence 61, Appl
593	5	2.6	145	3	US-08-630-172-5	Sequence 5, Appli	666	5	2.6	188	4	US-08-790-186A-1	Sequence 1, Appli
594	5	2.6	145	4	US-09-375-419-5	Sequence 5, Appli	667	5	2.6	188	4	US-08-861-269-7	Sequence 7, Appli
595	5	2.6	145	4	US-09-134-001C-5194	Sequence 5194, Ap	668	5	2.6	189	3	US-09-134-596-7	Sequence 7, Appli
596	5	2.6	145	4	US-09-656-450-12	Sequence 12, Appl	669	5	2.6	189	3	US-09-293-273-7	Sequence 7, Appli
597	5	2.6	146	4	US-09-209-525-2	Sequence 2, Appli	670	5	2.6	190	4	US-08-679-493A-184	Sequence 184, App
598	5	2.6	147	2	US-08-771-201-11	Sequence 11, Appl	671	5	2.6	191	2	US-08-290-665A-198	Sequence 198, App
599	5	2.6	147	3	US-08-946-329A-64	Sequence 64, Appl	672	5	2.6	191	2	US-08-290-665A-199	Sequence 199, App
600	5	2.6	147	4	US-09-134-001C-4542	Sequence 4542, Ap	673	5	2.6	191	2	US-08-290-665A-200	Sequence 200, App
601	5	2.6	147	4	US-09-134-001C-5518	Sequence 5518, Ap	674	5	2.6	191	2	US-08-290-665A-201	Sequence 201, App
602	5	2.6	148	3	US-08-946-329A-60	Sequence 60, Appl	675	5	2.6	191	2	US-08-290-665A-202	Sequence 202, App
603	5	2.6	148	3	US-08-946-329A-63	Sequence 63, Appl	676	5	2.6	191	2	US-08-290-665A-203	Sequence 203, App
604	5	2.6	148	3	US-08-946-329A-65	Sequence 65, Appl	677	5	2.6	191	2	US-08-290-665A-204	Sequence 204, App
605	5	2.6	148	4	US-09-352-990-4	Sequence 4, Appli	678	5	2.6	191	5	PCT-US95-10398-198	Sequence 198, App
606	5	2.6	148	4	US-09-171-461-18	Sequence 18, Appl	679	5	2.6	191	5	PCT-US95-10398-199	Sequence 199, App
607	5	2.6	149	3	US-08-584-031-17	Sequence 17, Appl	680	5	2.6	191	5	PCT-US95-10398-200	Sequence 200, App
608	5	2.6	150	4	US-08-862-124-5	Sequence 5, Appli	681	5	2.6	191	5	PCT-US95-10398-201	Sequence 201, App
609	5	2.6	150	4	US-09-134-001C-5014	Sequence 5014, Ap	682	5	2.6	191	5	PCT-US95-10398-202	Sequence 202, App
610	5	2.6	151	4	US-08-858-207A-298	Sequence 298, Appl	683	5	2.6	191	5	PCT-US95-10398-203	Sequence 203, App
611	5	2.6	155	1	US-08-530-010-12	Sequence 12, Appl	684	5	2.6	191	5	PCT-US95-10398-204	Sequence 204, App









GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:59:03 ; Search time 15 seconds  
(without alignments)  
376.613 Million cell updates/sec

Title: US-09-868-352-23  
Perfect score: 192  
Sequence: 1 MRKEVTPMLNKNYKPCQPF.....EEQDKEMTSAKQHLFVRKN 192

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	3.6	132	4	US-09-134-001C-4894
2	7	3.6	220	4	US-09-297-937C-7
3	7	3.6	284	3	US-09-053-197A-2
4	7	3.6	284	4	US-09-085-761A-2
5	7	3.6	317	4	US-08-937-067-6
6	7	3.6	1234	2	US-08-317-310A-15
7	7	3.6	1234	5	PCT-US95-13041-15
8	7	3.6	1243	2	US-08-537-139-2
9	7	3.6	1276	4	US-09-297-937C-13
10	7	3.6	1321	2	US-08-317-310A-16
11	7	3.6	1321	2	US-08-317-310A-64
12	7	3.6	1321	5	PCT-US95-13041-16
13	6	3.1	24	6	5240706-21
14	6	3.1	25	4	US-09-177-249-217
15	6	3.1	38	1	US-08-118-270-213
16	6	3.1	38	5	PCT-US93-08528-213
17	6	3.1	101	2	US-08-710-749-6
18	6	3.1	105	3	US-08-407-165-3
19	6	3.1	108	2	US-08-710-749-22
20	6	3.1	108	2	US-08-710-749-23
21	6	3.1	108	2	US-08-710-749-24
22	6	3.1	108	2	US-08-710-749-25
23	6	3.1	108	2	US-08-710-749-26
24	6	3.1	119	1	US-08-680-726A-62
25	6	3.1	119	4	US-09-092-409-62
26	6	3.1	139	1	US-08-680-726A-66
27	6	3.1	139	4	US-09-092-409-66

28	6	3.1	149	1	US-08-602-010A-18	Sequence 18, Appl
29	6	3.1	149	1	US-08-680-726A-18	Sequence 18, Appl
30	6	3.1	149	4	US-09-092-409-18	Sequence 18, Appl
31	6	3.1	196	4	US-08-975-762-32	Sequence 32, Appl
32	6	3.1	196	4	US-08-975-762-58	Sequence 58, Appl
33	6	3.1	196	4	US-08-821-324-32	Sequence 32, Appl
34	6	3.1	196	4	US-09-295-028-32	Sequence 32, Appl
35	6	3.1	196	4	US-09-295-028-58	Sequence 58, Appl
36	6	3.1	196	4	US-09-106-582-32	Sequence 32, Appl
37	6	3.1	196	4	US-09-106-582-58	Sequence 58, Appl
38	6	3.1	198	4	US-09-134-001C-3459	Sequence 3459, Ap
39	6	3.1	244	4	US-09-134-001C-4784	Sequence 4784, Ap
40	6	3.1	271	1	US-08-252-9950-11	Sequence 11, Appl
41	6	3.1	271	2	US-08-834-108-11	Sequence 11, Appl
42	6	3.1	294	4	US-09-424-349A-3	Sequence 3, Appl
43	6	3.1	331	4	US-09-457-046B-59	Sequence 59, Appl
44	6	3.1	369	2	US-08-663-566A-4	Sequence 4, Appl
45	6	3.1	369	2	US-08-023-610-4	Sequence 4, Appl
46	6	3.1	369	2	US-08-288-065A-4	Sequence 4, Appl
47	6	3.1	369	2	US-08-362-240A-4	Sequence 4, Appl
48	6	3.1	369	4	US-08-804-372A-2	Sequence 2, Appl
49	6	3.1	369	5	PCT-US95-10245-4	Sequence 4, Appl
50	6	3.1	385	2	US-08-892-715-2	Sequence 2, Appl
51	6	3.1	385	2	US-09-145-947-2	Sequence 2, Appl
52	6	3.1	385	4	US-09-265-642-2	Sequence 2, Appl
53	6	3.1	410	1	US-08-471-033-40	Sequence 40, Appl
54	6	3.1	410	1	US-08-471-033-43	Sequence 43, Appl
55	6	3.1	410	2	US-08-471-044-40	Sequence 40, Appl
56	6	3.1	410	2	US-08-471-044-43	Sequence 43, Appl
57	6	3.1	410	2	US-08-463-483A-40	Sequence 40, Appl
58	6	3.1	410	2	US-08-463-483A-43	Sequence 43, Appl
59	6	3.1	410	2	US-08-471-046A-40	Sequence 40, Appl
60	6	3.1	410	2	US-08-471-046A-43	Sequence 43, Appl
61	6	3.1	410	2	US-08-470-566B-40	Sequence 40, Appl
62	6	3.1	410	2	US-08-470-566B-43	Sequence 43, Appl
63	6	3.1	410	2	US-08-469-334-40	Sequence 40, Appl
64	6	3.1	410	2	US-08-469-334-43	Sequence 43, Appl
65	6	3.1	410	3	US-09-300-529-40	Sequence 40, Appl
66	6	3.1	410	3	US-09-300-529-43	Sequence 43, Appl
67	6	3.1	413	4	US-08-942-572-2	Sequence 2, Appl
68	6	3.1	446	4	US-08-960-780-52	Sequence 52, Appl
69	6	3.1	446	4	US-09-073-898-52	Sequence 52, Appl
70	6	3.1	449	1	US-08-471-033-46	Sequence 46, Appl
71	6	3.1	449	2	US-08-471-044-46	Sequence 46, Appl
72	6	3.1	449	2	US-08-463-483A-46	Sequence 46, Appl
73	6	3.1	449	2	US-08-471-046A-46	Sequence 46, Appl
74	6	3.1	449	2	US-08-470-566B-46	Sequence 46, Appl
75	6	3.1	449	2	US-08-469-334-46	Sequence 46, Appl
76	6	3.1	449	3	US-09-300-529-46	Sequence 46, Appl
77	6	3.1	462	1	US-08-471-033-2	Sequence 2, Appl
78	6	3.1	462	2	US-08-471-044-2	Sequence 2, Appl
79	6	3.1	462	2	US-08-463-483A-2	Sequence 2, Appl
80	6	3.1	462	2	US-08-471-046A-2	Sequence 2, Appl
81	6	3.1	462	2	US-08-470-566B-2	Sequence 2, Appl
82	6	3.1	462	2	US-08-469-334-2	Sequence 2, Appl
83	6	3.1	462	3	US-09-300-529-2	Sequence 2, Appl
84	6	3.1	472	4	US-09-088-425-1	Sequence 1, Appl
85	6	3.1	482	4	US-09-134-001C-4309	Sequence 4309, Ap
86	6	3.1	516	2	US-08-676-166A-7	Sequence 7, Appl
87	6	3.1	520	4	US-09-326-203A-2	Sequence 2, Appl
88	6	3.1	525	4	US-08-676-166A-3	Sequence 3, Appl
89	6	3.1	600	6	5240706-1	Patent No. 5240706
90	6	3.1	603	4	US-09-198-122-2	Sequence 2, Appl
91	6	3.1	607	4	US-09-204-208A-11	Sequence 11, Appl
92	6	3.1	621	1	US-08-295-814E-6	Sequence 6, Appl
93	6	3.1	621	5	PCT-US93-01959-6	Sequence 6, Appl
94	6	3.1	630	4	US-09-228-986-71	Sequence 71, Appl
95	6	3.1	650	4	US-09-232-191-29	Sequence 29, Appl
96	6	3.1	650	4	US-09-232-200-79	Sequence 79, Appl
97	6	3.1	650	4	US-09-232-200-97	Sequence 97, Appl
98	6	3.1	650	4	US-09-232-197-79	Sequence 79, Appl
99	6	3.1	650	4	US-09-232-197-97	Sequence 97, Appl
100	6	3.1	650	4	US-09-232-197-97	Sequence 97, Appl

THIS PAGE BLANK (USPTO)













CC with altered levels of polypeptide. The polypeptides are also useful for  
CC identifying agents (agonists and antagonists) that bind to them. Cells  
CC expressing the proteins are useful for identifying a therapeutic agent  
CC for use in treatment of a pathology related to aberrant expression or  
CC physiological interactions of the polypeptide. Vectors comprising  
CC the nucleic acids encoding the polypeptides and cells genetically  
CC engineered to express them are also useful for producing the proteins.  
CC The proteins are useful in genetic vaccination, testing and  
CC therapy, and can be used as nutritional supplements. They may be used to  
CC increase stem cell proliferation; to regulate haematopoiesis; and in  
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
CC in treatment of leukemias. AA029510-AA03304 represent the amino acid  
CC sequences of novel human secreted proteins of the invention.

XX Sequence 583 AA;

Query Match 9.3%; Score 93.5; DB 22; Length 583;  
Best Local Similarity 21.3%; Pred. No. 0.69;  
Matches 51; Conservative 41; Mismatches 64; Indels 83; Gaps 13;

QY 17 GPQIFHFNIVKSDDTFQVIVNEKSAFDVTVGQRFSEILLYK-----DFIVGDWGN 70

Db 92 GYGFVEDDLROADDAYVEL--NGKD-----LCGER---VIVEHARGPRD---GSYGSG 138

QY 71 Q-----LRLRGFYKDASTIRKNSRI-----SR-----LEDYK----- 98

Db 139 RSGYGYRRSGRDKYGPPTRTEDRLIVENLTSCSWQDLKDYMRQAGEVTYADAHKGRQKM 198

QY 99 -----EYCNFCGAYFVLE-----NPNRDIKFDDEP-----HKR----- 128

Db 199 KGVIEVSYSDMKRALEKLDGTEVNGKRLVLEDKPGSRRRRSYSRSHSRSRSHS 258

QY 129 RKSRSKSQSKSQTNRNRSOSNANAHFTSK--KRKDTKRQERHIKEODKEMTSKQ 185

Db 259 RKSRSGSGSKSHSRSRSGSRSGSRSGSRSGSRSGSRSGSRSGSRSGSRSH 317

RESULT 31

ABG10283

ID ABG10283 standard; Protein; 737 AA.

XX

AC ABG10283;

DT 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #10274.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS74470.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 20; SEQ ID No 40642; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 737 AA;

Query Match 9.2%; Score 93; DB 22; Length 737;

Best Local Similarity 26.1%; Pred. No. 1.1;

Matches 31; Conservative 25; Mismatches 27; Indels 36; Gaps 6;

QY 85 RKNRSIRSLIEDYKEYCNFCGAYFVLEPNRDIKFDDEP-----HKRRKSRKSOS 137

Db 400 RKNEKEKKVKDH-----KSNSEKDIRNSEKEDKYKNKKRAKSRKSRKS 446

QY 138 -SKSQTRNRSOSNANAHFTSKR---KDKTKRQERHIKEQ-----DKREMTSAKQ 184

Db 447 KEKSKSKERDSKHNRN-----EEKRMFRKELKGRDHENVKEKEKQSDSKGDKQERSRSE 501

RESULT 32

AAR13320

ID AAR13320 standard; Protein; 982 AA.

XX

AC AAR13320;

DT 17-DEC-2001 (updated)

DT 22-OCT-1991 (first entry)

XX Murine Natural Killer receptor.

DE NK; cytotoxic drugs; tumour cell.

XX Homo sapiens.

OS Key

EH Protein

FT /label= mature\_NK\_receptor

FT Peptide

FT /label= partial\_signal\_peptide

XX USN7535206-N.

PN 09-JUL-1991.

PD 08-JUN-1990; 90US-0143578.

PF 08-JUN-1990; 90US-0535206.

XX (USSH ) NAT INST OF HEALTH.

PA Ortaldo J, Young H, Anderson S;

PI

XX



DT 22-OCT-2001 (first entry)  
XX Human polypeptide SEQ ID NO 3439.  
XX  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI: 2001-442253/47.  
DR N-PSDB; AA159450.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX Example 6; SEQ ID NO 3439; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AA38642-AA42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
XX Sequence: 1150 AA;  
SQ

Query Match 9.5%; Score 95.5; DB 22; Length 1150;  
Best Local Similarity: 18.6%; Pred. No. 1;  
Matches 36; Conservative 39; Mismatches 64; Indels 55; Gaps 7;  
7;

QY 32 IEFQVINEKSAFDVAVFGORFSEILLKYDFIVGDWGNQDLRLRGYFKDASTI----- 84  
DB 753 LDYRLDNEKHSFVSLFAELFMELQOR-----DFG---VRI---YKSLSLPEKEDKK 800  
QY 85 -----RKNRIRSLRLEDYIKFCYCNFGCAYEVLENPNRPNRDKFDDERPHK-----R 128  
DB 801 KKKKSKDKDKKEDRDE-----TDEPKPKRKSGDDKDKEDRDKEDK 850

QY 129 RKRSKSSQSSKSGQTRNNRSQSNANAHFTSKK-----RKDTRRROERHIKESQDKE 178  
DB 851 RKDDSKDDDEEDNNQDEYDPMEEAEDEEDDRDEEMTKRDRDKNRYCKERPSPKD 910  
QY 179 MTSAKOHLFLVRKN 192  
DB 911 KEKEKTQMITINRD 924  
RESULT 28  
AAB40574  
ID AAB40574 standard; Protein; 2518 AA.  
XX  
XX AAB40574;  
XX  
XX 08-FEB-2001 (first entry)  
XX  
XX Human ORFX ORF338 polypeptide sequence SEQ ID NO:676.  
DE  
XX  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antichyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
XX Homo sapiens.  
XX  
XX WO200058473-A2.  
XX  
XX 05-OCT-2000.  
XX  
XX 31-MAR-2000; 2000WO-US08621.  
XX  
XX 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shinkets RA, Leach M;  
PI WPI: 2000-602362/57.  
DR N-PSDB; AAC74783.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 11; Page 784-790; 5507pp; English.  
XX  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antichyroid; antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,

```
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI: 2001-235357/24.
DR N-PSDB; AAH34397.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11: Page 7273-7275; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 540 AA;
Query Match 9.5%; Score 95.5; DB 22; Length 540;
Best Local Similarity 18.6%; Pred. No. 0.39;
Matches 36; Conservative 39; Mismatches 64; Indels 55; Gaps 7;
QY 32 IEFOLVINEKSAFDVTYFGQRFSEILLKYDFIVGDWNEQLRLRGFYKDASTI----- 84
Db 143 LDYRLDNKHSFEVSLFAELFNEMLQR-----DFG---VRI---YKSLLSLPEKEDK 190
QY 85 -----RKNRSIRLEDYIKCYNFCAYFVLEPNPNRDIKFDDEPHK-----R 128
Db 191 EKDKKSKDKERKKERDDE-----TDEPKPKRKSGDDKDKEDRDERKEDK 240
QY 129 RKSRSKSSQSKSQTRNNRSQSNANAHTSKK-----RKDKRQRERHIKEEQDK 178
Db 241 RKDSKDDDETEEDNQDEYDPMEAEAEDEDDRDEEEMTKRDDKRDINRYCKERPSKD 300
QY 179 MTSAKQHLLFVRKN 192
Db 301 KEKEKTOMITINRD 314
RESULT 26
AAB92973
ID AAB92973 standard; Protein; 718 AA.
AC AAB92973;
XX
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:11681.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX
```

```
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isojari T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyaama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 11681; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 718 AA;
Query Match 9.5%; Score 95.5; DB 22; Length 718;
Best Local Similarity 18.6%; Pred. No. 0.56;
Matches 36; Conservative 39; Mismatches 64; Indels 55; Gaps 7;
QY 32 IEFOLVINEKSAFDVTYFGQRFSEILLKYDFIVGDWNEQLRLRGFYKDASTI----- 84
Db 321 LDYRLDNKHSFEVSLFAELFNEMLQR-----DFG---VRI---YKSLLSLPEKEDK 368
QY 85 -----RKNRSIRLEDYIKCYNFCAYFVLEPNPNRDIKFDDEPHK-----R 128
Db 369 EKDKKSKDKERKKERDDE-----TDEPKPKRKSGDDKDKEDRDERKEDK 418
QY 129 RKSRSKSSQSKSQTRNNRSQSNANAHTSKK-----RKDKRQRERHIKEEQDK 178
Db 419 RKDSKDDDETEEDNQDEYDPMEAEAEDEDDRDEEEMTKRDDKRDINRYCKERPSKD 478
QY 179 MTSAKQHLLFVRKN 192
Db 479 KEKEKTOMITINRD 492
RESULT 27
AAM40294
ID AAM40294 standard; Protein; 1150 AA.
XX
XX AAM40294;
XX
```









```
RESULT 18
ABB68960
ID ABB68960 standard; Protein; 336 AA.
XX
AC ABB68960;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 33672.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WI; 2001-656860/75.
DR N-PSDB; ABL13063.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 33672; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 336 AA;
Query Match 9.7%; Score 97.5; DB 22; Length 336;
Best Local Similarity 30.1%; Pred. No. 0.13;
Matches 22; Conservative 17; Mismatches 29; Indels 5; Gaps 1;
Qy 115 PRDIKFDDEPHKRRKRSR-----KSSQSKSQTNRNRSQSNANAHFTSKKKRDKRKRRER 169
Db 79 PRTQSVDRSSSRSSRRHKAKKRRRRTRSSRSRSSSRSSSHRRKKKKKKYKK 138
Qy 170 HIKEQDKEMTSA 182
Db 139 HKKSHRRRSSQS 151
RESULT 19
AAM39441
ID AAM39441 standard; Protein; 323 AA.
XX
AC AAM39441;
XX
DT 22-OCT-2001 (first entry)
XX
```

```
DE Human polypeptide SEQ ID NO 2586.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
WI; 2001-142253/47.
DR N-PSDB; AAI58599.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2586; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 323 AA;
Query Match 9.5%; Score 96; DB 22; Length 323;
Best Local Similarity 30.6%; Pred. No. 0.18;
Matches 26; Conservative 18; Mismatches 17; Indels 24; Gaps 3;
Qy 122 DERPIKRRKRSR-----KSSQSKSQTNRNRSQSNANAHFTSK----- 158
Db 168 DEKRIKDKKSRTPPSRYNASSRRSSSRSSRRSSRRSSRPTSKTIKRRSSRSPRR 227
Qy 159 KRKD'KRQER-HIKEODKEMTSA 182
Db 228 NKDKIKREKRDHISRRERERSTS 252
RESULT 20
```



XX Novel nucleic acids and peptides derived from open reading frame x,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease --  
XX  
PS Claim 11; Page 2282-2283; 5507pp; English.  
XX  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressive;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 334 AA;  
Query Match 9.8%; Score 99; DB 21; Length 334;  
Best Local Similarity 29.0%; Pred. No. 0.093;  
Matches 29; Conservative 17; Mismatches 22; Indels 32; Gaps 4;  
QY 115 PRDIKFDDEPH-----RRKRSKSSQS-----SKQTRNNRSQSNANA 153  
DB 57 PRSHYDRRRRRRSSSSSYGSRGKRSRSGRGKSYRVQSRKSRTRRRSRRLRS 116  
QY 154 HFTSKKR-----KDKRKQERHIKEQDKEMTSK 183  
DB 117 HSRSSSRSHRTRSRDRRERKGRD-KEKREKEDKKG 155  
RESULT 15  
AAB43954  
ID AAB43954 standard; Protein; 376 AA.  
AC AAB43954;  
XX  
XX 08-FEB-2001 (first entry)  
DE Human cancer associated protein sequence SEQ ID NO:1399.  
XX  
KW Human; cancer associated gene; cancer antigen; detection; cancer;  
KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;  
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;  
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;  
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
KW allergic reaction; graft versus host disease; organ rejection;  
KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
KW neurological disease; drug screening.  
OS  
OS Homo sapiens.  
XX  
PN WO200055350-A1.  
XX  
XX 21-SEP-2000.  
PD  
XX 08-MAR-2000; 2000WO-US05882.  
XX

PR 12-MAR-1999; 99US-0124270.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Rosen CA, Iuben SM;  
PT  
DR WPI; 2000-587533/55.  
DR N-PSDB; AAC'8163.  
XX  
PT Novel isolated nucleic acids comprising sequences encoding peptides  
PT useful for treating or diagnosing e.g. cancer -  
PS  
PS Claim 11; Page 2077-2079; 2352pp; English.  
XX  
CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC in AAB43398 to AAB44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vulnery; immunomodulator;  
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;  
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;  
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells; to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 376 AA;  
Query Match 9.8%; Score 99; DB 21; Length 376;  
Best Local Similarity 34.3%; Pred. No. 0.11;  
Matches 24; Conservative 17; Mismatches 21; Indels 8; Gaps 2;  
QY 122 DERPHIRKRSKSSQS-----SKQTRNNRSQSNANAHTSKRKDKT--RQERHIKE 173  
DB 132 DKKEKIRHSRSRSRRRRTPSSRRHRSRRSRSSRSRRSRSSRSRRSRSSRSRE 191  
QY 174 EQDKEMTSK 183  
DB 192 RGRRSR;TSK 201  
RESULT 16  
AAE24596  
ID AAE24596 standard; Protein; 754 AA.  
XX  
XX AAE24596;  
XX  
XX 04-OCT-2002 (first entry)  
XX Human SR-cyp protein.  
XX  
XX Human; antipense; SR-cyp; Cdk-associated RS cyclophilin; inflammation;  
KW hyperproliferative disorder; cancer; prophylaxis; infection; therapy;  
KW tumour; CARP-cyp.  
XX  
XX Homo sapiens.  
XX  
XX WO200236809-A2.  
XX  
XX 10-MAY-2002  
XX  
XX 30-OCT-2001; 2001WO-US47335.  
XX

Db 354 -VOIIMDIKKAKKNNKKKQKNNKKK-----EKENKKKE 403  
QY 169 RHIEQDKREMTSAK 184  
Db 404 KNKKKEKSKKKE 419  
RESULT 13  
AAE15257  
ID AAE15257 standard; Protein; 624 AA.  
XX AAE15257;  
AC AAE15257;  
XX  
DT 07-MAR-2002 (first entry)  
XX  
DE Human RNA metabolism protein-20 (RMEP-20).  
XX  
XX Human; RNA metabolism protein-20; RMEP-20; gout; nervous system disorder;  
KW autoimmune; inflammatory; cell proliferative; developmental; thyroiditis;  
KW gene therapy; epilepsy; dementia; stroke; Alzheimer's disease; amnesia;  
KW Parkinson's disease; prion disease; insomnia; endocrine disorder; AIDS;  
KW Acquired Immune Deficiency Syndrome; mental disorder; allergy; anaemia;  
KW asthma; atherosclerosis; Crohn's disease; rheumatoid arthritis; vaccine;  
KW glomerulonephritis; multiple sclerosis; diabetes mellitus; osteoporosis;  
KW cancer; cirrhosis; hepatitis; psoriasis; transgenic animal; antiulcer;  
KW tranquiliser; drug screening; pancreatitis; renal tubular acidosis;  
KW systemic lupus erythematosus; colitis.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Domain 21..93  
FT /label= RRM  
FT /note= "RNA recognition motif"  
FT Domain 184..253  
FT /label= RRM  
FT /note= "RNA recognition motif"  
XX  
XX WO200183524-A2.  
XX  
XX 08-NOV-2001.  
XX  
XX 27-APR-2001; 2001WO-US13862.  
XX  
XX 28-APR-2000; 2000US-200184P.  
XX 04-MAY-2000; 2000US-201875P.  
XX 04-MAY-2000; 2000US-202050P.  
XX 06-JUN-2000; 2000US-210232P.  
XX 25-JUL-2000; 2000US-220553P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J, Hillman JL;  
XX Baughn MR, Yao MG, Burford N, Batra S, Policky JJ;  
XX  
XX WPI; 2002-034502/04.  
XX N-PSDB; AAD24393.  
XX  
XX New human RNA metabolism protein for diagnosing or treating nervous  
XX system disorders, autoimmune/inflammatory disorders, cell proliferative  
XX disorders and developmental disorders -  
XX  
XX Claim 1; Page 146-147; 196pp; English.  
XX  
XX The invention relates to human RNA metabolism proteins (RMEP) and their  
XX corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,  
XX treating and preventing nervous system disorders (epilepsy, dementia,  
XX stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);  
XX prion diseases; fatal familial insomnia, nutritional and metabolic  
XX diseases of the nervous system; inherited, metabolic, endocrine and  
XX toxic myopathy; mental disorders (mood, anxiety, schizophrenic disorders)  
XX anaemia and Tourette's disorder; autoimmune/inflammatory disorders (AIDS-

CC acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,  
CC atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,  
CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
CC arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,  
CC ulcerative colitis, and infections); cell proliferative disorders (cancer  
CC arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental  
CC disorders (renal tubular acidosis). RMEP DNA is useful in drug screening  
CC techniques, gene therapy and for creating transgenic animals. The present  
XX sequence is human RMEP-20 protein.  
SQ Sequence 624 AA;  
Query Match 10.0%; Score 101; DB 23; Length 624;  
Best Local Similarity 27.4%; Pred. No. 0.13;  
Matches 43; Conservative 26; Mismatches 68; Indels 20; Gaps 6;  
QY 32 IEF--QLVINEKSAFDVTFVGQRFSEILLKYDFIVGWGNEQLRGLRGFYKDASTIRKNSR 89  
Db 227 VEFADONSVPRALAFNGVFGDRPLKINHNNNAIVRP--PEMTPOAAAKELEEVKRR-- 281  
QY 90 ISKLEDYIKCYNFCGAYFVLEPNPRDIFDDERPHKRKSKSKSSQSSQTSNNRSQS 149  
Db 282 -----VREAQSFISAAIEPESGKSNRKGGRSGHTRSKRS---SSKSHSRKRSQS 331  
QY 150 --NANAHFTSKKR-KDTRKQERHKEEQDKEMTSAK 183  
Db 332 KHRSRHNRSRQKORRRKSKSPHKRKRKRERRKSR 368  
RESULT 14  
AAB41766  
ID AAB41766 standard; Protein; 334 AA.  
XX AAB41766;  
XX  
XX 08-FEB-2001 (first entry)  
XX  
XX Human ORFX ORF1530 polypeptide sequence SEQ ID NO:3060.  
XX  
XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
XX Homo sapiens.  
XX  
XX WO200058473-A2.  
XX  
XX 05-OCT-2000.  
XX  
XX 31-MAR-2000; 2000WO-US08621.  
XX  
XX 31-MAR-1999; 99US-0127607.  
XX 02-APR-1999; 99US-0127636.  
XX 05-APR-1999; 99US-0127728.  
XX 30-MAR-2000; 2000US-0540763.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shimkets RA, Leach M;  
XX  
XX WPI; 2000-602362/57.  
XX N-PSDB; AAC75975.

```
XX SQ Sequence 604 AA:
Query Match 10.4%; Score 105; DB 22; Length 604;
Best Local Similarity 22.1%; Pred. No. 0.048;
Matches 36; Conservative 34; Mismatches 65; Indels 28; Gaps 5;

Qy 46 VTFVGFQSEILLKYD-FTVGDWGNOLRLRGFYKDASTIRKNS-----RISRLDLY 96
Db 218 LTOLKESFOOVLEDERSEYHLKGERARWQQRMRKMSQEICTLKKEKODMRVVELEERS 277
Qy 97 IKEYCNFGCAYEVLEN-----PNPRDIKFDDEPHKRRKRSKSSQSKSQTNRNR--S 147
Db 278 LSK-----LKNQMAEPLPPEPPAVPSEVLEQLHLRKELEVERAGELQSQVKNNQHS 327
Qy 148 QSNANAHFTSKRKDTKRQRERHIKEQDKEMTSAKOHLFLFVR 190
Db 328 LLNRQEEIRIQEERLRKQERLQEQHEKRLQALAKPHCAFSR 370

RESULT 11
AAV22206
ID AAV22206 standard; Protein; 361 AA.
XX AC AAV22206;
XX DT 13-SEP-1999 (first entry)
XX DE Biorhythm marker protein.
XX KW Biorhythm marker gene; circadian rhythm.
XX OS Rattus sp.
XX PN JP11169185-A.
XX PD 29-JUN-1999.
XX PF 12-DEC-1997; 97JP-0362890.
XX PR 12-DEC-1997; 97JP-0362890.
XX PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
XX PS (SAKA ) OTSUKA PHARM CO LTD.
XX DR WPI; 1999-422623/36.
XX PT N-PSDB; AAX84569.
XX PT Structure and function of a new biorhythm marker gene - and protein
XX PT encoded by it
XX PS Claim 7; Page 14-16; 23pp; Japanese.
XX CC This sequence is the biorhythm marker protein of the invention. The
XX CC protein is also referred to as a circadian rhythm protein.
XX SQ Sequence 361 AA;
Query Match 10.2%; Score 102.5; DB 20; Length 361;
Best Local Similarity 33.0%; Pred. No. 0.045;
Matches 30; Conservative 15; Mismatches 33; Indels 13; Gaps 4;

Qy 103 FGCA-----YFVLENPNPRDIKFDDEPHKRRKRSKSSQSKSQTNRNRSSQ--NANA 153
Db 35 FNCSELTQSGFFLPESGSKNERKGRSRSHTRSKRS---SSKSHSRKRSKRSQSKRSRS 91
Qy 154 HFTSKKR-KDTKRRQERHIKEQDKEMTSK 183
Db 92 HNRSRQKDRRSKSPHKRSKSRERKSR 122

RESULT 12
AAB18187
```

```
ID XX AAB18187 standard; Protein; 508 AA.
XX AC AAB18187;
XX DT 07-NOV-2000 (first entry)
XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:44.
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX KW antimalarial; malaria; protozoacide; infection; insecticide.
XX OS Plasmodium falciparum.
XX PN WO200025728-A2.
XX PD 11-MAY-2000.
XX PF 05-NOV-1999; 99WO-US26796.
XX PR 05-NOV-1998; 98US-0107131.
XX PA (HOFF/) HOFFMAN S.
XX PA (CARU/) CARUCCI D.
XX PA (GARD/) GARDNER M.
XX PA (VENT/) VENTER J C.
XX PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX DR WPI; 2000-355347/31.
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,
XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the
XX PT diagnosis of P.falciparum infection -
XX PS Disclosure: Page 108-110; 577pp; English.
XX CC The present invention describes proteins and their fragments (I) encoded
XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX CC vaccines against P. falciparum infection comprising (I) or (II).
XX CC (I) and (II) are useful for the development of vaccines against
XX CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX CC antibody raised to immunogens comprising the sequences of (I), are
XX CC useful in the detection of infection with P. falciparum. Furthermore,
XX CC (I) especially when they are rifins or secreted or membrane proteins)
XX CC can aid the identification of drugs to treat or prevent P. falciparum
XX CC infection, or they can be used to identify drug resistance in
XX CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX CC subsequent identification of proteins encoded by it will help to expand
XX CC our understanding of parasite biology, a process hampered by the
XX CC complexity of the parasitic lifecycle, and provide new targets for
XX CC vaccine and drug development. Parasite resistance to drugs and mosquito
XX CC resistance to insecticides have led to a resurgence of malaria in many
XX CC parts of the world, and there is a pressing need for vaccines and new
XX CC drugs. AAV20078 to AAV70287 and AAB18144 to AAB18352 represent nucleotide
XX CC and protein sequences given in the present invention, but which are not
XX CC specifically mentioned within the specification.
XX SQ Sequence 508 AA;
Query Match 10.0%; Score 101; DB 21; Length 508;
Best Local Similarity 20.9%; Pred. No. 0.099;
Matches 41; Conservative 38; Mismatches 61; Indels 56; Gaps 9;

Qy 5 VTPENLWYKYPGQFIHFENIV-----KSDDIE----FQVINEKSAFQVTVFG 50
Db 264 ITKE-CVWKYD-----DNLIHDKQHKREKKSNIENMISHIYEKEQSHDIC--- 312
Qy 51 QRFSBILKYDFIVGDWGNQOLRLRGFYKDASTIRKRSRLDLY--IKEYCNFGCAYF 108
Db 313 -----NVLEENKEEKYNNLQKDVITNCNDKV-KLEEYHHEKELNN----- 353
Qy 109 VLENPN?RDIKFDDEPHKRRKRSKSSQSKSQTNRNRSSQSNANAHFTSKRKDKTRQE 168
```

Matches	45;	Conservative	24;	Mismatches	29;	Indels	3;	Gaps	2
<hr/>									
QY	26	IVKSDDIEFQLVINEKS	AFVTVEGQRFSFELLKYDFIVGDWGNQGLRLRGFYKDASTIR	85					
		: :   :	: :   :	: :   :	: :   :	: :   :	: :   :	:	:
Dd	1	MIKVQQVFELIEEYRECFDEIFESARYSDILDKDYDVVGDCYDQDLRLKGFYKDSNKA	60						
<hr/>									
QY	86	K-NSRISLRLEDYIKCYCNFCGCAYFVLENPNPRDI--KFDD	123						
		: :   :	: :   :	: :   :	: :   :	: :   :	: :   :	:	:
Dd	61	EISKRFSSIQDYILEYCNCFGCPYFVRRRLSPNEFIEIDDK	101						
<hr/>									
RESULT 9									
ABBS0067	ID	ABBS0067 standard; Protein; 91 AA.							
XX	AC	ABBS0067;							
XX	DT	05-FEB-2002 (first entry)							
DE	L	Listeria monocytogenes protein #271.							
XX	Antibacterial;	gene therapy; vaccine; biosynthesis; biodegradation;							
KW	vitamin B12;	bacterial infection; disease.							
KW	Listeria monocytogenes.								
OS	WO200177335-A2.								
XX	18-OCT-2001.								
XX	11-APR-2001;	2001WO-FR01118.							
XX	11-APR-2000;	2000FR-0004629.							
PR	(INSP )	INST PASTEUR.							
PA	Buchrieser C,	Frangoul L, Couve E, Rusniok C, Fsihi H, Dehoux P;							
PI	Dussurget O,	Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;							
PI	Danilels J,	Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;							
PI	Dominguez-Bernal G,	Garrido-Garcia-P, Tierrez-Martinez A, Amend A;							
PI	Chakraborty T,	Domann E, Hain T, Berche P, Charbit A, Durant L;							
PI	Perez-diaz J,	Baquero F, Garcia Del Portillo F, Gomez-Lopez N;							
PI	Madueno E,	De Pablos B, Wehlant J, Kaerst U, Entian K, Hauf J;							
PI	Rose M,	Voss H;							
XX	WPI:	2002-010914/01.							
DR	Genomic sequence for	Listeria monocytogenes, useful e.g. for treatment							
PT	and prevention of	Listeria and related bacterial infections, and							
PT	related polypeptides								
XX	Claim 6:	SEQ ID No 2772; 192bp; French.							
XX	The present invention relates to the genome sequence of	Listeria							
CC	monocytogenes EGD-e (see ABA03041).	The genome sequence and fragments of							
CC	it are useful for selecting probes and primers for detecting genes in	L.							
CC	polymorphisms and other genomes.	The present sequence is a protein							
CC	encoded by the genome sequence of the present invention.	Proteins							
CC	expressed from the genome sequence are useful for raising specific	antibodies, identification of L. monocytogenes and related organisms, and							
CC	for biosynthesis and biodegradation, especially biosynthesis of Vitamin	B12.							
CC	The genome sequence and proteins encoded by it are also useful for	selecting compounds that regulate gene expression and cell replication							
CC	and modulate L. monocytogenes-related diseases.	In addition, the genome							
CC	sequence and proteins encoded by it are useful in pharmaceutical and	vaccines compositions for the treatment or prevention of infections by L.							
CC	monocytogenes and related organisms.								
CC	Note: The sequence data for this patent did not form part of the printed	specification, but was obtained in electronic format directly from WIPO							
CC	at ftp.wipo.int/pub/published_pct_sequences.								
XX	Sequence	91 AA;							
SQ									

**Query Match** 23.7%; Score 239.5; DB 23; Length 91;  
**Best Local Similarity** 51.2%; Pred. No. 7e-17;  
**Matches** 43; Conservative 22; Mismatches 14; Indels 5; Gaps  
0

QY	31	DIEFQLVINEKSAFDVTVFEGORFESEILLKYDFIVGDWNGNEOLRLRGFYKDASTIRK--N 87
		: ::::   :     :   :     :    :    :    :    :    :    :    :    :
Db	7	LDAYEITTYRDAFOEEKLNRFSDILGRYDVIWGDMGYDLRLKLGFFEDDN--RRAAYD 64
		::   L:::        ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   :
QY	88	SRTSLRELDYIKEYNFCGCAIFYLE 111
		:   L:::        ::   ::   ::   ::   ::   ::   ::   ::   ::   :
Db	65	NKISTLKEIYYEVCNCAYFIK 88
		:

**RESULT 10**

ID ABGI1566 standard; Protein; 604 AA.

XX AC ABGI1566;

XX XX

DT DT

TT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #15157.

DE DE

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

KW KW

OS Homo sapiens.

QS QS

PX WC200175067-A2.

PN PN

PD PD

PP PP

PF 30-MAR-2001; 2001WO-US08631.

XX XX

PR 31-MAR-2000; 2000US-0540217.

PR PR

PT 23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

PA PA

XX Drmanac RT, Liu C, Tang YT;

PI PI

XX WP1; 2001-639362/73.

DR DR

NR N-PSDB; AAS79353.

XX XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

PS PS

SS Claim 20; SEQ ID NO 45525; 103pp; English.

XX XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromoso

CC CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence t

CC for identifying expressed genes. (I) is useful in gene therapy techn

CC to restore normal activity of (II) or to treat disease states involv

CC CC (II). (II) is useful for generating antibodies against it, detectin

CC quantitating a polypeptide in tissue, as molecular weight markers an

CC a food supplement. (II) and its binding partners are useful in medic

CC CC imaging of sites expressing (II). (I) and (II) are useful for treati

CC disorders involving aberrant protein expression or biological activit

CC The polypeptide and polynucleotide sequences have applications in

CC CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiver

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from W

CC at ftp.wipo.int/pub/published\_pct\_sequences.

OC OC

AA	Sequence	91 AA;
SQ		

```
SQ Sequence 251 AA;
Query Match 32.9%; Score 331.5; DB 23; Length 251;
Best Local Similarity 34.4%; Pred. No. 9.6e-26;
Matches 75; Conservative 38; Mismatches 64; Indels 41; Gaps 5;

Qy 1 MRKEVTPMLNKNYPGQFIHFENIVKSDIEFOLVINERKSAFDVTVFGORFSEILLY 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MAKVIDESKLNKNYPGEHMAVGVEVQVQGTTFHIVHNYREAFDAEKLQRFSDVLSKY 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 61 DFIVGDMGNEQLRLRGFYKDASTIRK----NSRISRLLEDYKCYNFCGCAVFLVLENPNR 116
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 DYIVGDMGFEQLRLKGF--STSRKMLADNKIDHLEDYVNEYCNYGCAVFLVLRIRTK 117
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 117 DIKFDEDER-----PHKRRKRSKSSQSSKSSQTRNNRSQSNAHAFTSKKSKD 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 118 DEAFVSEKLFTEKELKQGFDPKRRKRNRNRNWARDQKVTREDKRSEN-----SSEARKD 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 163 TKRROERH-----IKEEQDKEMTSAKQH 185
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 173 FKIREKSTDRPKPVTDNKNKVSYSKSSOERKTDNKKQN 210
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
ABP40049
ID ABP40049 standard; Protein; 132 AA.
XX
AC ABP40049;
XX
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4894.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
PL Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
DR N-PSDB; ABN92594.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 4894; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 132 AA;
Query Match 24.1%; Score 243.5; DB 22; Length 129;
Best Local Similarity 44.6%; Pred. No. 4.3e-17;

Query Match 25.2%; Score 254.5; DB 23; Length 132;
Best Local Similarity 45.6%; Pred. No. 3.3e-18;
Matches 47; Conservative 24; Mismatches 29; Indels 3; Gaps 2;

Qy 24 ENIVKSDIEFOLVINERKSAFDVTVFGORFSEILLYKDYFIVGDMGNEQLRLRGFYKDAST 83
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 ENMIVKJQOYFELIEFYRECDEFETFSARYSDILDYDVVVGDTGYDGLRLKGFYKDSNK 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 84 IRK-NSRISRLLEDYKCYNFCGCAVFLVLENPNRDI--KEDDE 123
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 KAEISKRPSSIQDKILEYCNFCGCPYFVVRRULSPNEFIEIDDK 104
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
AAG82114
ID AAG82114 standard; Protein; 129 AA.
XX
AC AAG82114;
XX
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:1322.
XX
KW Staphylococcus epidermidis SRL strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX
XX Kimmerly WJ;
XX
XX WPI; 2001-316495/33.
DR N-PSDB; AAH52964.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 376-377; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 129 AA;
Query Match 24.1%; Score 243.5; DB 22; Length 129;
Best Local Similarity 44.6%; Pred. No. 4.3e-17;
```





XX 02-JUL-2002 (first entry)  
DT Streptococcus polypeptide SEQ ID NO 150.  
DE  
DE Streptococcus GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX Streptococcus pyogenes.  
OS  
XX WO200234771-A2.  
PN  
XX 02-MAY-2002.  
PD  
XX 29-OCT-2001; 2001WO-GB04789.  
PF  
XX 27-OCT-2000; 2000GB-0026333.  
PR 24-NOV-2000; 2000GB-0028727.  
PR 07-MAR-2001; 2001GB-0005640.  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
PA  
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
PI Tettelin H;  
PI  
XX WPI: 2002-352536/38.  
DR N-PSDB: ABN66118.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection  
PT or disease caused by Streptococcus bacteria, such as meningitis, and  
PT for detecting a compound that binds to the protein -  
XX  
XX Claim 1; Page 3168; 4525pp; English.  
PS  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) are used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.  
XX  
SQ Sequence 214 AA;  
Query Match 55.4%; Score 558.5; DB 23; Length 214;  
Best Local Similarity 51.7%; Pred. No. 4.3e-49;  
Matches 107; Conservative 42; Mismatches 41; Indels 17; Gaps 4;  
Oy 1 MRKEVTPMLNKNYPGQFTHFENIVKSDDDIEFQVINEKSAFDVTVFGORFSEILLKY 60  
Db 9 MKKELSPMYNKNYPGPKFTHFEQVRAEGIDLLLEDVKNADFDTTFGQRYTEVLKY 68  
Oy 61 DFIVGDCNEQLRLKGFYKSDASTIRKNSRISLEDEYIKFCNFCAYFVLENPNRDIKF 120  
Db 69 DYIVGDCNEQLRLKGFYKSDDKIKTKRISLEDEYIKFCNFCAYFVLENLHPQDIKF 128  
Oy 121 DDERPHKRKS-RSKSQ-----SSKSQTNNRSQSNANAHFTSKRRKDKTR 165  
Db 129 EEPORPRKKSFKSKSRNKNPNYSNOQPATPKSKS-RASKEQEPENQAFTSKRRSNTK 187  
Oy 166 RQERHKEEQDKEMTSAKOHLLFVRKN 192

Db 188 HKEKS-KRNOTSQLNTKISHFIIRKDD 213  
RESULT 4  
AAU37821  
ID AAU37821 standard; Protein; 176 AA.  
XX  
AC AAU37821;  
XX  
DT 14-FEB-2002 (first entry)  
XX  
XX Streptococcus pneumoniae cellular proliferation protein #250.  
DE  
XX  
KW Antisense: prokaryotic cellular proliferation protein;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Streptococcus pneumoniae.  
XX WO200170955-A2.  
PN  
XX 27-SEP-2001.  
PD  
XX  
XX 21-MAR-2001; 2001WO-US09180.  
PF  
XX  
XX 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Haselbeck H, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RJ, Xu HH;  
PI  
XX WPI: 2001-611495/70.  
DR N-PSDB: AA55680.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Example 3; Seq ID No 13414; 511pp; English.  
PS  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 176 AA;  
Query Match 44.8%; Score 452.5; DB 22; Length 176;  
Best Local Similarity 47.6%; Pred. No. 2.4e-38;  
Matches 91; Conservative 35; Mismatches 50; Indels 15; Gaps 4;  
Oy 1 MRKEVTPMLNKNYPGQFTHFENIVKSDDDIEFQVINEKSAFDVTVFGORFSEILLKY 60

PR 04-JAN-1999; 99GB-0000084.  
 PR 04-JAN-1999; 99GB-0000085.  
 PR 04-JAN-1999; 99GB-0000086.  
 PR 28-JAN-1999; 99GB-0001916.  
 PR 28-JAN-1999; 99GB-0001922.  
 XX  
 PA (MICR-) MICROSCIENCE LTD.  
 XX  
 PI Hughes MJG, Santangelo JD, Lane JD, Everest P, Feldman R;  
 PI Moore JC, Wilson RK, Dobson RJ, Dougan G;  
 XX  
 DR WPI: 2000-442674/38.  
 DR N-PSDB; AAA51368.  
 XX  
 XX New peptides useful for treatment and prevention of conditions  
 PT associated with Streptococcal infection are obtained from group B  
 PT Streptococcus  
 XX  
 PS Claim 2: Page 47-48; 63pp; English.  
 XX  
 CC The S. agalactiae pho3-1 protein has homology with proteins of unknown  
 CC function in S. pyogenes, S. pneumoniae, Bacillus subtilis (yutD) and  
 CC Enterococcus faecalis. The B. subtilis yutD gene is located in a  
 CC chromosomal region containing genes involved in cell wall synthesis.  
 CC Peptides derived from group B Streptococcus (Streptococcus agalactiae)  
 CC proteins encoded by genes pho1-13, pho3-21, pho2-15, pho3-18, pho3-22,  
 CC pho3-3, pho3-17, pho2-5, pho1-5, pho3-1, pho3-23, pho3-50, pho1-14,  
 CC pho2-10, pho3-14, pho3-24 and pho3-29. The peptides are useful for  
 CC screening potential drugs, or for the detection of virulence, and for the  
 CC manufacture of a medicament for use in the treatment or prevention of  
 CC infections such as focal infection (including osteomyelitis, septic  
 CC arthritis, abscesses and endophthalmitis) and urinary tract infections  
 CC caused by group B Streptococcus. The peptides and vaccines comprising the  
 CC peptides are useful in treatment of chronic mastitis, especially in cows  
 CC (i.e. for veterinary purposes).  
 XX  
 SQ Sequence 192 AA;  
 Query Match 100.0%; Score 1009; DB 21; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-95;  
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRKEVTPMLNKNYPGQFIHFENIVKSDDEIFQLVINEKSAFDTVFGQRFSEILLKY 60  
 DB 1 MRKEVTPMLNKNYPGQFIHFENIVKSDDEIFQLVINEKSAFDTVFGQRFSEILLKY 60  
 QY 61 DFIVGDWGNELRLRGFYKDASTIRKNSRISRLDYIKCYCNFGCAFYVLENPNPRDIKF 120  
 DB 61 DFIVGDWGNELRLRGFYKDASTIRKNSRISRLDYIKCYCNFGCAFYVLENPNPRDIKF 120  
 QY 121 DDERPHKRRKRSKSSKSSQTRNNRSQSNANAHTSKKRKDTKRQERHIKEEQDKEMT 180  
 DB 121 DDERPHKRRKRSKSSKSSQTRNNRSQSNANAHTSKKRKDTKRQERHIKEEQDKEMT 180  
 QY 181 SAKQHLLEFVRKN 192  
 DB 181 SAKQHLLEFVRKN 192  
 RESULT 2  
 ABP25486  
 ID ABP25486 standard; Protein: 192 AA.  
 XX  
 AC ABP25486;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 148.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX

OS Streptococcus agalactiae.  
 XX WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 XX 29-OCT-2001; 2001WO-GB04789.  
 XX  
 XX 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 DR WPI: 2002-352536/38.  
 DR N-PSDB; ABN66117.  
 XX  
 CC New Streptococcus protein for the treatment or prevention of infection  
 CC or disease caused by Streptococcus bacteria, such as meningitis, and  
 CC for detecting a compound that binds to the protein -  
 PS Claim 1: Page 3168; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX  
 SQ Sequence 192 AA;  
 Query Match 97.1%; Score 980; DB 23; Length 192;  
 Best Local Similarity 97.4%; Pred. No. 2.4e-92;  
 Matches 186; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MRKEVTPMLNKNYPGQFIHFENIVKSDDEIFQLVINEKSAFDTVFGQRFSEILLKY 60  
 DB 1 MRKEVTPMLNKNYPGQFIHFENIVKSDDEIFQLVINEKSAFDTVFGQRFSEILLKY 60  
 QY 61 DFIVGDWGNELRLRGFYKDASTIRKNSRISRLDYIKCYCNFGCAFYVLENPNPRDIKF 120  
 DB 61 DFIVGDWGNELRLRGFYKDASTIRKNSRISRLDYIKCYCNFGCAFYVLENPNPRDIKF 120  
 QY 121 DDERPHKRRKRSKSSKSSQTRNNRSQSNANAHTSKKRKDTKRQERHIKEEQDKEMT 180  
 DB 121 DDERPHKRRKRSKSSKSSQTRNNRSQSNANAHTSKKRKDTKRQERHIKEEQDKEMT 180  
 QY 181 SAKQHLLEFVRKN 191  
 DB 181 SAKQHFVIRKK 191  
 RESULT 3  
 ABP25487  
 ID ABP25487 standard; Protein: 214 AA.  
 XX  
 AC ABP25487;





Db 2152 ----ENQRHQRDRRSHRASERSLGRYTDVDTGLGTLDSMTTQSGDLPSPKRDQGRPK 2208

QY 162 DTKRQERH 170

Db 2209 DRKHROHH 2217

RESULT 39

US-09-134-001C-3930

; Sequence 3930, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3930

; LENGTH: 564

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3930

Query Match

Best Local Similarity 7.7%; Score 77.5; DB 4; Length 564;

Matches 44; Conservative 36; Mismatches 60; Indels 57; Gaps 9;

QY 11 NTNKPQPOFIHFENIVKSDDIIEFQLVINEKSAFDVTVFQGRFSEILLKYDFIVGDWNE 70

Db 152 NANRY---TFVWKKSI---QNHESRMNENSKALYHELVIKTIPEIKKDHD---NDLTKE 202

QY 71 QLRLGRFYKD-----ASTIRKNSRISRLIEDYIKEYCNFCAYFVLENPNR 116

Db 203 EDLIGSHLDKEIDLNQHDNEKCTKIRKQIRLKRTK--IKKYK----- 246

QY 117 DIKFDEPRKRRKRSKSSQSKSOTRNNRSQSNANAHTSKRKDKTKRQERHIKEOD 176

Db 247 --QIND---YSQRKHKEVQKSIKDRNSYSKTDHDTAF-----MRMKEDHMKNGOL 293

QY 177 K-----EMTSARQHLLF 188

Db 294 KPGYNLQIATNSQFVLF 310

RESULT 40

US-09-356-952-5

; Sequence 5, Application US/09356952

; Patent No. 6117663

; GENERAL INFORMATION:

; APPLICANT: Boriack-Stodin, Ann

; APPLICANT: Margarit, S. M.

; APPLICANT: Bor-Sogli, Dafna

; APPLICANT: Cole, Philip

; APPLICANT: Kuriyan, John

; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: 600-1-228N

; CURRENT APPLICATION NUMBER: US/09/356,952

; CURRENT FILING DATE: 1999-07-19

; EARLIER APPLICATION NUMBER: 60/093,631

; EARLIER FILING DATE: 1998-07-21

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn ver. 2.0

; SEQ ID NO 5

; LENGTH: 1048

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-356-952-5

Query Match

Best Local Similarity 7.7%; Score 77.5; DB 3; Length 1048;

Matches 43; Conservative 25; Mismatches 66; Indels 47; Gaps 7;

QY 8 EMLNKNYVPGPOFIHFENIVKSDDIIEFQLVINEKSAFDVTVFQGRFSEILL---KYDFI 63

Db 872 QSLNKLMDPKKNFIINRYNELKS-----LHSAPCVPFEGVYLSDLTFTTDSGNPDYL 921

QY 64 VGDWGNQLRLRGFYKDASTIRKNSRISRLIEDYIKEYCNF-----GCAYEV 109

Db 922 VLEHG-----LKGVDHKKYINFNKR-SRLVDILQELIYFKKTHYDFTKDRTVIECISNS 975

QY 110 LENPNPQDIKFD-----DERPHKRRKRSKSSQSKSOTRNNRSQSNANAHTSKRKDKTK 164

Db 976 LENIPHIKQYQLSLIIEFKPRKKVVPVNSNS-----NNKSQEKSRDDQDTDEGKTSTK 1027

QY 165 R 165

Db 1028 K 1028

Search completed: February 26, 2003, 09:55:08

Job time : 25 secs



```
QY 169 --RHIKEE 174
      | |||
Db 65 LTRGAKEE 72

RESULT 35
US-09-203-453-5
; Sequence 5, Application US/09203453
; Patent No. 6426411
; GENERAL INFORMATION:
; APPLICANT: Spiegelman, Bruce M.; Puigserver, Pere; Wu, Zhidan and Adelmant, Guillaume
; TITLE OF INVENTION: PGC-1, A NOVEL BROWN FAT PPAR{SYMBOL 103 \f "symbol"} COACTIVATOR
; FILE REFERENCE: DFN-023CP
; CURRENT APPLICATION NUMBER: US/09/203,453
; CURRENT FILING DATE: 1998-12-01
; EARLIER APPLICATION NUMBER: 09/086,912
; EARLIER FILING DATE: 1998-05-29
; EARLIER APPLICATION NUMBER: 60/048,107
; EARLIER FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 5
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-203-453-5

      Query Match      7.8%; Score 79; DB 4; Length 798;
      Best Local Similarity 21.1%; Pred. No. 4.7;
      Matches 38; Conservative 29; Mismatches 49; Indels 64; Gaps 8;

QY 66 DWGNEQ-----LRLRGFYKDASTIRKNSRISRLDYIKYCNF-----GCAYF-- 108
Db 488 DFNESQFKLPMFINSGLAMDGLFDSDD--KSKLSPYWDGTQSYSLFNVSPPSCSFSNS 545

QY 109 -VLENPFRDIFKDDERPHKR-----KSRKSQSKSQTRN----- 144
Db 546 PCRDVSPPKSLF-SQRQMRSRSRFSRHRSCSRSPYSRSRSPGSRSSRSRSCYYIE 604

QY 145 -----NRSQSNANAHFTSKKRDKTRR-----QERHIKEQDKWMTSAKQ 184
Db 605 SSHYRHRTHRNSPLYSRSPYSRPRDYSEYEQHERLKRREYRREYKRESERAKQ 664

RESULT 36
US-09-041-886-21
; Sequence 21, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
```

```
;
;
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1182 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-21

      Query Match      7.8%; Score 78.5; DB 4; Length 1182;
      Best Local Similarity 25.08; Pred. No. 9.1;
      Matches 28; Conservative 17; Mismatches 30; Indels 37; Gaps 5;

QY 80 DASTIFKNSRI-----SRLEDYIKYCNFGCAYFVLENPFRDIFKDDERPHKRKRSK 134
      | : : : : | : : : | : : : | : : : | : : : | : : : |
Db 794 DTSPMKRSASVLGPKARLDY-----SLERVPE-----ENQRHHQRDRSH 837

QY 135 SQSSKQTR-----NRSQSNANAHFTSKKR-----KDTKRQERH 170
      | : | : | : | : | : | : | : | : | : | : | : |
Db 838 RASERLGRYTDVDTGLTDLMTTQSGDLPSKRDQGRPKDKRKHQHHH 889

RESULT 37
US-08-149-097D-6
; Sequence 36, Application US/08149097D
; Patent No. 584236
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: Feldman, Daniel
; APPLICANT: McCue, Ann
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/149,097D
; FILING DATE: 05-NOV-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/105,536
; FILING DATE: 11-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US92/06903
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,231
; FILING DATE: 13-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/868,354
; FILING DATE: 10-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,206
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,250
```



```
; APPLICANT: He, Tong-Chuan
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
; TITLE OF INVENTION: Prevent Cancer
; FILE REFERENCE: 1107 75741
; CURRENT APPLICATION NUMBER: US/09/136,605A
; CURRENT FILING DATE: 1998-08-20
; EARLIER APPLICATION NUMBER: 08/821,355
; EARLIER FILING DATE: 1997-03-20
; EARLIER APPLICATION NUMBER: 09/003,687
; EARLIER FILING DATE: 1998-01-06
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2973
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-136-605-7

Query Match      7.9%; Score 79.5; DB 4; Length 2973;
Best Local Similarity 23.5%; Pred. No. 25;
Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps 3;

QY 78 YKDASTIRKNSRISRLDYIKYCNFCGAYFVLENPNRPDIKEDDERPH-----126
Db 1135 YEDDKPTNTYSERYSEEQHEE-----ERPTNYSIKYNEERKHVDQPIDYSLKY 1183
QY 127 -----KRRKSRKSSQSSKSTQNNRSQSNANAHFTSKRKDKTKRQERHIKEQDK 177
Db 1184 ATDIPSSQKQSFSSKSSQSSQSKTEHMSSEN---TSTPSSNAKRONLHPSSAQSR 1239

RESULT 33
US-08-698-407-4
; Sequence 4, Application US/08698407
; Patent No. 5856128
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/698,407
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0116 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
```

```
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 267188
US-08-698-407-4

Query Match      7.8%; Score 79; DB 2; Length 475;
Best Local Similarity 33.8%; Pred. No. 2.3;
Matches 23; Conservative 12; Mismatches 21; Indels 12; Gaps 3;

QY 111 ENPNPRDIK-FDDERPHKRRKRSKSSQSSKSTQNNRSQSNANAHFTSKRKDKTKRQE- 168
Db 13 ENKQERDKENRHRKRSRSDRKRKRSRDRNRDQRSAS-----RDRRRRSKP 64
QY 169 --RHIKEE 174
Db 65 LTRGAKEE 72

RESULT 34
US-09-195-855-4
; Sequence 4, Application US/09195855
; Patent No. 6015788
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,855
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/698,407
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0116 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 267188
US-09-195-855-4

Query Match      7.8%; Score 79; DB 3; Length 475;
Best Local Similarity 33.8%; Pred. No. 2.3;
Matches 23; Conservative 12; Mismatches 21; Indels 12; Gaps 3;

QY 111 ENPNPRDIK-FDDERPHKRRKRSKSSQSSKSTQNNRSQSNANAHFTSKRKDKTKRQE- 168
Db 13 ENKQERDKENRHRKRSRSDRKRKRSRDRNRDQRSAS-----RDRRRRSKP 64
```



SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-450-582-2

Query Match 7.9%; Score 79.5; DB 3; Length 2843;  
Best Local Similarity 23.5%; Pred. No. 23;  
Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps 3;  
QY 78 YKDASTIRKNSRISRLDYIKIYCNFCGAYFVLENNPRDIKFDDEPH----- 126  
DB 1135 YEDDKPTNYSERYSEEEH-----ERPTNYSIKYNEKRHVDPIDYSLKY 1183  
QY 127 -----KRRKSRKSSQSKSQTNRNRSOSNANAHTSKRKDTKRRQERHIKEQDK 177  
DB 1184 ATDIPSSQKQSFSSKSSQSSQSKTEHSSSEN---TTPSSNAKRQNLHPSSAQSR 1239

RESULT 28  
US-08-450-582-7  
Sequence 7, Application US/08450582  
Patent No. 6114124

GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE

THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA

ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,582  
FILING DATE:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/452,655  
FILING DATE: 25-MAY-1995  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
US-08-450-582-7

Query Match 7.9%; Score 79.5; DB 3; Length 2843;  
Best Local Similarity 23.5%; Pred. No. 23;  
Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps 3;  
QY 78 YKDASTIRKNSRISRLDYIKIYCNFCGAYFVLENNPRDIKFDDEPH----- 126  
DB 1135 YEDDKPTNYSERYSEEEH-----ERPTNYSIKYNEKRHVDPIDYSLKY 1183  
QY 127 -----KRRKSRKSSQSKSQTNRNRSOSNANAHTSKRKDTKRRQERHIKEQDK 177  
DB 1184 ATDIPSSQKQSFSSKSSQSSQSKTEHSSSEN---TTPSSNAKRQNLHPSSAQSR 1239

RESULT 29  
US-08-449-731-2  
Sequence 2, Application US/08449731  
Patent No. 6413727

GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE

THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti, LTD  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA

ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,731  
FILING DATE: 25-May-1995

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/289,548  
FILING DATE: 12-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.46943

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-08-449-731-2

COUNTRY: US  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/370,235A  
FILING DATE: 01-JAN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.48688  
TELECOMMUNICATION INFORMATION:

; INFORMATION FOR SEQ ID NO: 2







```

; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-540-824-25

Query Match          7.9%; Score 80; DB 4; Length 810;
Best Local Similarity 21.4%; Pred. No. 3.7;
Matches 34; Conservative 35; Mismatches 54; Indels 36; Gaps 8;

QY 36 LVINEKSAFDVTVFG-----QRFSEILLKYDFIVGDWGNQELRLRGFYKADSTIRKNSKI 90
Db 635 VLLNVWESLDATISTVWKSPLAFNEIILPIQOLLSAYTS---KYSDFEKPRLNKNVLEKL 691
QY 91 SLREYDIKEYCNFGCAYFVLEPNPRDI-----KFD-----DERPHKRRKRSKSSOSSKS 140
Db 692 TKFTEHIP-----LALQNHKPVSIPTHAPKYEEFNPNPKSKSYDPDRTREINKMKA 742
QY 141 QTRNNRSOSNANAHTSKK-RKDTRRQERHIKEQDKE 178
Db 743 QLKKEK-----FTMKERIRKDAFEARQRI-EENKE 773

RESULT 17
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07741.940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC

US-09-540-824-25
Query Match          7.9%; Score 80; DB 4; Length 810;
Best Local Similarity 21.4%; Pred. No. 3.7;
Matches 34; Conservative 35; Mismatches 54; Indels 36; Gaps 8;

QY 36 LVINEKSAFDVTVFG-----QRFSEILLKYDFIVGDWGNQELRLRGFYKADSTIRKNSKI 90
Db 635 VLLNVWESLDATISTVWKSPLAFNEIILPIQOLLSAYTS---KYSDFEKPRLNKNVLEKL 691
QY 91 SLREYDIKEYCNFGCAYFVLEPNPRDI-----KFD-----DERPHKRRKRSKSSOSSKS 140
Db 692 TKFTEHIP-----LALQNHKPVSIPTHAPKYEEFNPNPKSKSYDPDRTREINKMKA 742
QY 141 QTRNNRSOSNANAHTSKK-RKDTRRQERHIKEQDKE 178
Db 743 QLKKEK-----FTMKERIRKDAFEARQRI-EENKE 773

RESULT 17
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07741.940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC

US-07-741-940-7
Query Match          7.9%; Score 79.5; DB 1; Length 2842;
Best Local Similarity 23.5%; Pred. No. 23;
Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps 3;

QY 78 YKDASTIRKNSRISLEIDYIKEYCNFGCAYFVLEPNPNRDIKFDDEPH--4-----126
Db 1134 YEDDRFTNYSERYSEEEH--4-----EPTNYSIKYNEKRHVDPIDYSLKY 1182
QY 127 -----KRRKRSKSSKSSQTRNNRSQSNANAHFTSKKRKDTKRRQERHIKEEQDK 177
Db 1183 ATDIPESQKQSFSEKSSGSSQSSKTEHMSSEN---TSTPSSNAKRONQLHPSSAQR 1238

RESULT 18
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-08-289-548A-7
Query Match          7.9%; Score 79.5; DB 1; Length 2842;
Best Local Similarity 23.5%; Pred. No. 23;
Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps 3;
```





Db 33 DDIPLAKSRKRKRVESDYED-----EDEVPLKRRKLSNGRAKK 71

Qy 136 QSSKSOTNNRSQSNANAHFTSKRKDKTKRQER-HIKEODKEMTSAKQ 184

Db 72 Q-VKTETKVKRKPESANKS-KSTSKKDKTKVKKETTVKKESKATSTKVKE 119

RESULT 11

US-08-973-831-2

Sequence 2, Application US/08973831

Patent No. 5942386

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric B.

APPLICANT: Gerhold, David L.

APPLICANT: Strauss, Allyson Cole

TITLE OF INVENTION: Anti-fungal Agents and Methods of

TITLE OF INVENTION: Identifying and Using the Same

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &

ADDRESSEE: No. 5942386ris

STREET: One Liberty Place, 46th floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/973.831

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/485,621

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: DeLuca, Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-1584

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 780 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-973-831-2

Query Match 8.1%; Score 82; DB 2; Length 780;

Best Local Similarity 28.2%; Pred. No. 2.2;

Matches 31; Conservative 19; Mismatches 32; Indels 28; Gaps 6;

Qy 80 DASTIRKNSRISRLP-DYIKYCNFGCAYFVLEPNPRDIKFDDEPHKRRK---SRSKS 135

Db 33 DDIPLAKSRKRKRVESDYED-----EDEVPLKRRKLSNGRAKK 71

Qy 136 QSSKSOTNNRSQSNANAHFTSKRKDKTKRQER-HIKEODKEMTSAKQ 184

Db 72 Q-VKTETKVKRKPESANKS-KSTSKKDKTKVKKETTVKKESKATSTKVKE 119

RESULT 12

PCT-US96-09530A-2

Sequence 2, Application PC/TUS9609530A

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric B.

APPLICANT: Gerhold, David L.

APPLICANT: Strauss, Allyson Cole

TITLE OF INVENTION: Anti-fungal Agents and Methods of

TITLE OF INVENTION: Identifying and Using the Same

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris

STREET: One Liberty Place, 46th floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/09530A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/485,621

FILING DATE: 07-JUN-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/000,399

FILING DATE: 21-JUN-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: DeLuca, Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-1970

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 780 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US96-09530A-2

Query Match 8.1%; Score 82; DB 5; Length 780;

Best Local Similarity 28.2%; Pred. No. 2.2;

Matches 31; Conservative 19; Mismatches 32; Indels 28; Gaps 6;

Qy 80 DASTIRKNSRISRLP-DYIKYCNFGCAYFVLEPNPRDIKFDDEPHKRRK---SRSKS 135

Db 33 DDIPLAKSRKRKRVESDYED-----EDEVPLKRRKLSNGRAKK 71

Qy 136 QSSKSOTNNRSQSNANAHFTSKRKDKTKRQER-HIKEODKEMTSAKQ 184

Db 72 Q-VKTETKVKRKPESANKS-KSTSKKDKTKVKKETTVKKESKATSTKVKE 119

RESULT 13

US-09-086-912-2

Sequence 2, Application US/09086912

Patent No. 6156192

GENERAL INFORMATION:

APPLICANT: Bruce M. Spiegelman, Pere Puigserver and Zhidan Wu

TITLE OF INVENTION: PGC-1, A No. 6166192el Brown Fat PPAR[SYMBOL

TITLE OF INVENTION: 103 \f "Symbol"] Coactivator

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk



APPLICANT: YAN, Chunhua et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: CLO00758  
CURRENT APPLICATION NUMBER: US/09/810,671  
CURRENT FILING DATE: 2001-06-08  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 445  
TYPE: PRT  
ORGANISM: Human  
US-09-810-671-2

Query Match 8.2%; Score 83; DB 4; Length 445;  
Best Local Similarity 27.7%; Pred. No. 0.79;  
Matches 28; Conservative 17; Mismatches 44; Indels 12; Gaps 3;  
Qy 78 YKASTIRKNSRISLEDYIKCYCNFCAYFVLEPNPNRDIKFDDEPHKRRKRSKSSQS 137  
Db 20 YLEARSL--NERYDRRYVDEYRNDYCEGYV-----PRHYRDIESGYRHCKSSSVRS 72  
Qy 138 SKSOTRNNRSOSNANAHFTSKRKDKTKRRQERHIKEEQDK 178  
Db 73 RRSPPKRR-----NRHCSHQSRKSKSHRRKRSRSEDDEE 108

RESULT 6  
US-08-307-499-28  
Sequence 28, Application US/08307499  
Patent No. 5651972  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Vi uela, Eladio  
APPLICANT: Gibbs, E.P.J.  
TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a  
TITLE OF INVENTION: Live Vaccine Vector  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/307,499  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/908,241  
FILING DATE: 1-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/908,630  
FILING DATE: 29-JUN-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/342,212  
FILING DATE: 21-APR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: 0F35.1.FWCCI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 236 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-307-499-28  
Query Match 8.2%; Score 82.5; DB 1; Length 236;  
Best Local Similarity 24.3%; Pred. No. 0.38;  
Matches 35; Conservative 25; Mismatches 45; Indels 39; Gaps 8;  
Qy 5 VTPEMLN-----YNKYPGQFIHFENIVKSDDIEFOLV--INERSAFDVTF----- 49  
Db 79 VVPHMNVASIIYSEY---EYLLKSNVKNKKINTYLDKINKYHSIDDIIFMYLHWKK 135  
Qy 50 -----GQRSEILLKYD-----FIVGDMGNEQLRLRGFYKDASTIRKNSRISLEDYI 97  
Db 136 YNNTCAGCKLFKE-LMKYDILATKYIYNDIINT-----YKEGDTISINIRLCKCKDDII 187  
Qy 98 KEYCNFCAYFVLEPNPNRDIKFD 121  
Db 188 K-HCKSSIGMFAILLSSKIIDVDFD 210  
RESULT 7  
US-09-299-268-2E  
Sequence 28, Application US/09299268  
Patent No. 6217882  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Vi uela, Eladio  
APPLICANT: Gibbs, E.P.J.  
TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a  
TITLE OF INVENTION: Live Vaccine Vector  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/299,268  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/901,127  
FILING DATE:  
APPLICATION NUMBER: US 07/908,241  
FILING DATE: 1-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/908,630  
FILING DATE: 29-JUN-1992  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/342,212  
FILING DATE: 21-APR-1992  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: 0F35.1.FWCCI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 28:

```

; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914.999
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 601-1-078
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 732 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Dictyostellium discoideum
; US-08-914-999-8
;
; Query Match 8.6%; Score 87; DB 4; Length 732;
; Best Local Similarity 20.9%; Pred. No. 0.57;
; Matches 45; Conservative 29; Mismatches 69; Indels 72; Gaps 11;
;
QY 12 YNKYPG-----PQ-FIHENIVKSDDIEFQVLVINEKSAFDVTVCQRFSEILL-KY 60
DB 242 YNNNSGFVSDNERNTPQSFHP-----TYEHSNHOLLI-----IDTQGVGDHYTDPQIHTY 292
QY 61 D---FTVGDWNGEQLRLGRFYKDAFTIRKNSIRSLIEDYIKCYCNFGCAYFVLENPNPR- 116
DB 293 DGVGFQIGNLGO-----KGFELDTHK-----CNAICQYLNLSQINPKS 332
QY 117 -----DIKFDDERPHRRKRSKSSQSKSQTNRNRSQSNANAHFTSK-----R 160
DB 333 EKSDCGTVPRLIFDPTSDRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSISLSVEIS 392
QY 161 KDKRQERHKEEQ-----DKEMTSK 183
DB 393 SSKERNDRDSRQLFVSDNGTNTLNKRSKSK 427
;
RESULT 3
US-07-667-276A-4
; Sequence 4, Application US/07667276A
; Patent No. 5470971
; GENERAL INFORMATION:
; APPLICANT: Kondo, Keiji
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING
; TITLE OF INVENTION: THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welser & Associates
; STREET: 230 S. Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914.999
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 601-1-078
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 732 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Dictyostellium discoideum
; US-08-914-999-8
;
; Query Match 8.6%; Score 87; DB 4; Length 732;
; Best Local Similarity 20.9%; Pred. No. 0.57;
; Matches 45; Conservative 29; Mismatches 69; Indels 72; Gaps 11;
;
QY 12 YNKYPG-----PQ-FIHENIVKSDDIEFQVLVINEKSAFDVTVCQRFSEILL-KY 60
DB 242 YNNNSGFVSDNERNTPQSFHP-----TYEHSNHOLLI-----IDTQGVGDHYTDPQIHTY 292
QY 61 D---FTVGDWNGEQLRLGRFYKDAFTIRKNSIRSLIEDYIKCYCNFGCAYFVLENPNPR- 116
DB 293 DGVGFQIGNLGO-----KGFELDTHK-----CNAICQYLNLSQINPKS 332
QY 117 -----DIKFDDERPHRRKRSKSSQSKSQTNRNRSQSNANAHFTSK-----R 160
DB 333 EKSDCGTVPRLIFDPTSDRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSISLSVEIS 392
QY 161 KDKRQERHKEEQ-----DKEMTSK 183
DB 393 SSKERNDRDSRQLFVSDNGTNTLNKRSKSK 427
;
RESULT 3
US-07-667-276A-4
; Sequence 4, Application US/07667276A
; Patent No. 5470971
; GENERAL INFORMATION:
; APPLICANT: Kondo, Keiji
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING
; TITLE OF INVENTION: THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welser & Associates
; STREET: 230 S. Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914.999
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 601-1-078
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 732 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Dictyostellium discoideum
; US-08-914-999-8
;
; Query Match 8.6%; Score 87; DB 4; Length 732;
; Best Local Similarity 20.9%; Pred. No. 0.57;
; Matches 45; Conservative 29; Mismatches 69; Indels 72; Gaps 11;
;
QY 12 YNKYPG-----PQ-FIHENIVKSDDIEFQVLVINEKSAFDVTVCQRFSEILL-KY 60
DB 242 YNNNSGFVSDNERNTPQSFHP-----TYEHSNHOLLI-----IDTQGVGDHYTDPQIHTY 292
QY 61 D---FTVGDWNGEQLRLGRFYKDAFTIRKNSIRSLIEDYIKCYCNFGCAYFVLENPNPR- 116
DB 293 DGVGFQIGNLGO-----KGFELDTHK-----CNAICQYLNLSQINPKS 332
QY 117 -----DIKFDDERPHRRKRSKSSQSKSQTNRNRSQSNANAHFTSK-----R 160
DB 333 EKSDCGTVPRLIFDPTSDRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSISLSVEIS 392
QY 161 KDKRQERHKEEQ-----DKEMTSK 183
DB 393 SSKERNDRDSRQLFVSDNGTNTLNKRSKSK 427
;
RESULT 3
US-07-667-276A-4
; Sequence 4, Application US/07667276A
; Patent No. 5470971
; GENERAL INFORMATION:
; APPLICANT: Kondo, Keiji
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING
; TITLE OF INVENTION: THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welser & Associates
; STREET: 230 S. Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914.999
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 601-1-078
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 732 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Dictyostellium discoideum
; US-08-914-999-8
;
; Query Match 8.6%; Score 87; DB 4; Length 732;
; Best Local Similarity 20.9%; Pred. No. 0.57;
; Matches 45; Conservative 29; Mismatches 69; Indels 72; Gaps 11;
;
QY 12 YNKYPG-----PQ-FIHENIVKSDDIEFQVLVINEKSAFDVTVCQRFSEILL-KY 60
DB 242 YNNNSGFVSDNERNTPQSFHP-----TYEHSNHOLLI-----IDTQGVGDHYTDPQIHTY 292
QY 61 D---FTVGDWNGEQLRLGRFYKDAFTIRKNSIRSLIEDYIKCYCNFGCAYFVLENPNPR- 116
DB 293 DGVGFQIGNLGO-----KGFELDTHK-----CNAICQYLNLSQINPKS 332
QY 117 -----DIKFDDERPHRRKRSKSSQSKSQTNRNRSQSNANAHFTSK-----R 160
DB 333 EKSDCGTVPRLIFDPTSDRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSISLSVEIS 392
QY 161 KDKRQERHKEEQ-----DKEMTSK 183
DB 393 SSKERNDRDSRQLFVSDNGTNTLNKRSKSK 427
;
RESULT 3
US-07-667-276A-4
; Sequence 4, Application US/07667276A
; Patent No. 5470971
; GENERAL INFORMATION:
; APPLICANT: Kondo, Keiji
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING
; TITLE OF INVENTION: THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welser & Associates
; STREET: 230 S. Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914.999
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 601-1-078
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 732 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Dictyostellium discoideum
; US-08-914-999-8
;
; Query Match 8.6%; Score 87; DB 4; Length 732;
; Best Local Similarity 20.9%; Pred. No. 0.57;
; Matches 45; Conservative 29; Mismatches 69; Indels 72; Gaps 11;
;
QY 12 YNKYPG-----PQ-FIHENIVKSDDIEFQVLVINEKSAFDVTVCQRFSEILL-KY 60
DB 242 YNNNSGFVSDNERNTPQSFHP-----TYEHSNHOLLI-----IDTQGVGDHYTDP
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:51:52 ; Search time 15 seconds  
(without alignments)  
376.613 Million cell updates/sec

Title: US-09-868-352-23  
Perfect score: 1009  
Sequence: 1 MRKEVTPEMLNKNKPGPQF.....EEQDKEMTSKQHLLEVRKN 192

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	254.5	25.2	132	4	US-09-134-001C-4894
2	87	8.6	732	4	US-08-914-999-8
3	86	8.5	414	1	US-07-667-276A-4
4	83	8.2	427	4	US-09-810-671-4
5	83	8.2	445	4	US-09-810-671-2
6	82.5	8.2	236	1	US-08-307-499-28
7	82.5	8.2	236	4	US-09-299-268-28
8	82	8.1	429	4	US-09-810-671-5
9	82	8.1	484	4	US-09-457-040B-12
10	82	8.1	780	1	US-08-485-621-2
11	82	8.1	780	2	US-08-973-831-2
12	82	8.1	780	5	PCT-US96-09530A-2
13	82	8.1	797	4	US-09-086-912-2
14	82	8.1	797	4	US-09-203-453-2
15	80	7.9	262	4	US-09-134-001C-5532
16	80	7.9	810	4	US-09-540-824-25
17	79.5	7.9	2842	1	US-07-741-940-7
18	79.5	7.9	2842	1	US-08-289-548A-7
19	79.5	7.9	2842	1	US-08-452-654-7
20	79.5	7.9	2842	1	US-08-449-731-7
21	79.5	7.9	2843	1	US-07-741-940-2
22	79.5	7.9	2843	1	US-08-289-548A-2
23	79.5	7.9	2843	1	US-08-452-654-2
24	79.5	7.9	2843	1	US-08-452-655B-2
25	79.5	7.9	2843	1	US-08-452-655B-7
26	79.5	7.9	2843	2	US-08-370-235A-2
27	79.5	7.9	2843	3	US-08-450-582-2

28	79.5	7.9	2843	3	US-08-450-582-7	Sequence 7, Appli
29	79.5	7.9	2843	4	US-08-449-731-2	Sequence 2, Appli
30	79.5	7.9	2973	2	US-08-821-355A-7	Sequence 7, Appli
31	79.5	7.9	2973	2	US-09-003-687A-7	Sequence 7, Appli
32	79.5	7.9	2973	4	US-09-136-605-7	Sequence 7, Appli
33	79	7.8	475	2	US-08-698-407-4	Sequence 4, Appli
34	79	7.8	475	3	US-09-195-855-4	Sequence 5, Appli
35	79	7.8	798	4	US-09-203-453-5	Sequence 5, Appli
36	78.5	7.8	1182	4	US-09-041-886-21	Sequence 21, Appl
37	78.5	7.8	2265	2	US-08-149-097D-36	Sequence 36, Appl
38	78.5	7.8	2509	2	US-08-149-097D-35	Sequence 35, Appl
39	77.5	7.7	564	4	US-09-134-001C-3930	Sequence 3930, Ap
40	77.5	7.7	1048	3	US-09-356-952-5	Sequence 5, Appli
41	77.5	7.7	1898	1	US-08-056-200-94	Sequence 94, Appl
42	77.5	7.7	1898	2	US-08-800-644-94	Sequence 94, Appl
43	76.5	7.6	265	4	US-09-149-476-497	Sequence 497, App
44	76.5	7.6	216	1	US-08-185-424B-1	Sequence 1, Appli
45	76.5	7.6	432	2	US-08-933-750C-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1

US-09-134-001C-4894  
; Sequence 4894, Application US/09134001C  
; Patent No. 6300370  
; GENERAL INFORMATION:  
; APPLICANT: L'inn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4894  
; LENGTH: 132  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4894

Query Match 25.2%; Score 254.5; DB 4; Length 132;  
Best Local Similarity 45.6%; Pred. No. 4.9e-20;  
Matches 47; Conservative 24; Mismatches 29; Indels 3; Gaps 2;

Qy	24	ENIVKSDIEFQLVINEKSAFDVTVEGGRFSEILLKYDFIVGDWGNQOLRLRGFYKDAST 83
Db	2	ENMIVDOOYFELIEYRECFDEEIFSARYSDILDKYDVVGDIYDQRLAGFYRDSNK 61
Qy	84	IRK-NSRISRLIEDYIKYCNFCGAYFVLENNPRDI--KFDDE 123
Db	62	KAEIKRESSIQDYILEYCNFCGCPYFVVRLLSPNEEFIEIDDK 104

RESULT 2

US-08-914-999-8  
; Sequence 8, Application US/08914999  
; Patent No. 6346406  
; GENERAL INFORMATION:  
; APPLICANT: Ryzanov, Alexey G.  
; APPLICANT: Hait, William N.  
; APPLICANT: Pavur, Karen S.  
; TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)  
; TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: DAVID A. JACKSON, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; FLOOR









Db 36 GYGFVEFEDSRDADDAVYELNGKELCGEHVIVEHARGPRDRDGYSGSRGGGYSSRR 95  
QY 74 LRGFYKDASTIRKNSRI-----SR-----LEDYI-----KEYCNFGCAYP-- 108  
Db 96 TSGRDYKGPVTEYVRLVIVENUSRCSCWOOLKDFMRQAGEVTVYADAHKEKRTNEGVIENRS 155  
QY 109 -----VLE-----NPNPRDIKFDDEP-----HKRRKRSRKS-QSSKSO 141  
Db 156 YSDMKRALDKLDGTEINGNIRLIEDKPTSHRRSYSGSRSRSSRRSRSSRRSR 215

QY 142 TRN-NRSQSNANAHFTSKRKDKTKRQERHIKEEQDKEMTSAKQH 185  
Db 216 SRISKRSRSGRSRSGRSRSGKRSKRSKSPKSDRGSHSH 260

## RESULT 27

E70013  
Hypothetical protein yukC - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C:Accession: E70013  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni, C.; Bron, S.; Brouillette, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinols, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Winters, P.; Wipat, A.; Yamada, T.; Yano, K.; Yoshida, K.; Yata, K.; Yochiyama, K.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: E70013  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-451 <KUN>  
A:Cross-references: GB:299120; GB:AL009126; NID:g2635613; PIDN:CAB15177.1; PID:el184266;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yukC

Query Match 9.1%; Score 91.5; DB 2; Length 451;  
Best Local Similarity 21.9%; Pred. No. 6.1;  
Matches 44; Conservative 31; Mismatches 71; Indels 55; Gaps 9;

QY 4 EVTPMLNKNYKPGQFIHFENI--VKSDIEFQVLVINEKSAFDVTVFGQRFSEILLKYD 61  
Db 275 ESLPSVQYQL--ATSYVEVENLGSAAKTKNIENVTLOSDF-----QHFLYWI----- 321

QY 62 FTVGQWNEQLRLRGFYKDASTIRKNSRISRLDYIKEYCNFGCAFYVL-----EN 112  
Db 332 ---DYG-----RGEYKATSI---GRKLENDYI-----YFALAKYKQLLSED 359

QY 113 PNPRDI-----KFDDEPHKRRKRSKRSQSSKQSTNRNRSQSNANAHFTSKRKDKT 163  
Db 360 TNDEDIQKELDSVNSELEKAQKERQENQSNSETSLVDTSEQOTQDEKQAEKAAEEK 419

QY 164 KRQERHIKEEQDKEMTSAKQ 184  
Db 420 AAAEEKAKKEQKEDEKKE 440

## RESULT 28

A47328  
natural killer cell tumor-recognition protein - human  
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor  
C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 24-Sep-1999  
C:Accession: A47328  
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.  
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993  
A:Title: A cyclophilin-related protein involved in the function of natural killer cell  
A:Reference number: A47328; MUID:9313824; PMID:8421688  
A:Accession: A47328  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1403 <AND>  
A:Cross-references: NID:L04288; NID:gl81251; PIDN:AAA35734.1; PID:gl81252  
A:Experimental source: NK killer cells from adult blood  
A:Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIP:122800)  
C:Genetics:  
A:Gene: GDB:NKTF

A:Cross-references: GDB:137171; OMIM:161565

A:Map position: 3p23-3p21  
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology  
C:Keywords: alternative splicing; lymphocyte  
F:60-230/Domain: cyclophilin homology <CYP>

Query Match 9.1%; Score 91.5; DB 1; Length 1403;  
Best Local Similarity 28.1%; Pred. No. 21;  
Matches 32; Conservative 14; Mismatches 41; Indels 27; Gaps 4;

QY 78 YKD---ASTIRKNSRISRLDYIKEYCNFGCAFYVLENPDPRIKFDDEPHKRRKRSK 134  
Db 735 YSDRSKESPSRSGRSRSGRSYSRSTR-----SRSLASSHSR 774

QY 135 SQSSKSKQTRN---NRSQSNANAHFTSKRKDKTKRQERHIKEEQDKEMTSAKQH 185  
Db 775 SPSSK:HSRKNYSDHSQCRSSSYTIS-SDGRRAKRRLRSGGKNSVSHKKH 827

## RESULT 29

S57157  
Hypothetical protein YJRI34c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein J2120  
C:Species: Saccharomyces cerevisiae  
C:Date: 23-Aug-1995 #sequence\_revision 08-Sep-1995 #text\_change 19-Apr-2002  
C:Accession: S57157  
R:Rose, M.; Koetter, P.; Entian, K.D.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56848  
A:Accession: S57157  
A:Molecule type: DNA

A:Residues: 1-737 <ROS>  
A:Cross-references: EMBL:Z49634; NID:gl015869; PIDN:CAA89665.1; PID:gl015870; GSPDB:G

A:Gene: SGD:SGM1; MIPS:YJRI34C

A:Cross-references: SGD:S0003895

A:Map position: 10R

Query Match 9.0%; Score 90.5; DB 2; Length 707;  
Best Local Similarity 22.2%; Pred. No. 12;  
Matches 41; Conservative 33; Mismatches 52; Indels 59; Gaps 8;

QY 3 KEVTEMLNKNYKPGQFIHFENIYKSDIEFQVLVINEKSAFDVTVFGQRFSEILLKYDF 62  
Db 127 KEISFEYLRANK-----QIDLTNEL--NRKSQIETT--DSSFFKLKEKDD 169

QY 63 IVGDKGNEQLRLRGFYKDASTIRKNSRISRLDYIKEYCNFGCAFYVLENPDPRIKFD 122  
Db 170 LI---DQLRKEGAKLAETELRQSNQIKALTKVK-----DLEYE- 205

QY 123 ERPHARRKRSKRSKRSQSSKQSTNRNRSQS-----NANAHFTSK-----RKDKRRQERHI 171  
Db 206 -----VSELNDSSAQSVENNELQSLYHNITQGLAEATNKLKADKQKESLETLEKNI 258

QY 172 KEEQ 176  
Db 259 KEKD 263





```
Db 300 RGRSRSTSK 309
      : : : :
RESULT 17
A48133
pre-mRNA splicing SRP75 - human
C:Species: Homo sapiens (man)
C>Date: 30-Jun-1995 #sequence_revision 30-Jun-1995 #text_change 05-Nov-1999
C:Accession: A48133
Mol. Cell. Biol. 13, 4023-4028, 1993
R:Zahler, A.M.; Neugebauer, K.M.; Stolk, J.A.; Roth, M.B.
A:Title: Human SR proteins and isolation of a cDNA encoding SRP75.
A:Reference number: A48133; MUID:93309435; PMID:8321209
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-494 <ZAH>
A:Cross-references: GB:L14076; NID:g307437; PIDN:AAA36649.1; PID:g307438
A:Note: parts of this sequence were confirmed by peptide sequencing
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoproteins
C:Keywords: phosphoprotein; pre-mRNA splicing
F:3-62/Domain: ribonucleoprotein repeat homology <RRM3>
F:105-167/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 9.8%; Score 99; DB 2; Length 494;
Best Local Similarity 22.0%; Pred. No. 1.7; Mismatches 75; Indels 70; Gaps 10;
Matches 51; Conservative 36;

QY 17 GPQIFHEINVKSDIEFQVLVINEKSAFDVTV-----FGQRFSG----- 54
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 36 GYGVEFDLRLDADDAVYELNGDLCCERVIVEHARGPRDGSVGSRGSGYGRSRGRDK 95
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 55 ---EILKYDFIVGD-----WGN--EQLRLRG--FYKDASTIRKNRISRLDYIKEYC 101
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 96 YGPPTREYRLIVENLSRCSQDLKDYMQAGEVTYADAHKGRKN-----EGVIEFV 148
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 102 NFGCAYFVLE-----NPNPRDKFDDERP-----HKR-----RKSRKS 135
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 149 SYSDMKRALEKLDGTENVGRKIRLVEDKPGRRRRYSRSRSHSRSRSHSRSHSRSR 208
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 136 QSSKSQTRNNRSQSNANAHFTSK--KKDKTKRQERHIKEEDKEMTSAKOH 185
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 209 GSSKSHSKSRSRSGSRSRSRSRSRSRSRSRSRSRSRSPKDKSRSRSH 260
      | : : : : | : : : : | : : : : | : : : : | : : : : |

RESULT 18
T37189
hypothetical protein C02H7.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Feb-2000
C:Accession: T37189
R:Leimbac, D.; Minx, M.
A:Description: The sequence of C. elegans cosmid C02H7.
A:Reference number: Z20523
A:Accession: T37189
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-535 <LEI>
A:Cross-references: EMBL:U49945; PIDN:AAC47924.1; GSPDB:GN00029; CESP:C02H7.1
A:Experimental source: strain Bristol N2; clone C02H7
C:Genetics:
A:Gene: CESP:C02H7.1
A:Map position: X
A:Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 9.8%; Score 99; DB 2; Length 535;
Best Local Similarity 19.0%; Pred. No. 1.9;
Matches 40; Conservative 39; Mismatches 87; Indels 44; Gaps 5;

QY 1 MRKEVTPMLNKNYPGPQFTHFENIVKSDDIEFQVLVINEKSAFDVTVFGQRFSEILLKY 60
      | | | | | : : : : | : : : : | : : : : | : : : : | : : : : |
```

```
Db 1 MSVEETREIL-----EKVIQKPOLQLDQLSRPPFKFIVDI-----VSNVIKST 43
QY 61 DFIVGWGNEQLRLRGFYKDASTIRKNSRISRLD-----YIK----- 98
      : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 44 GYLKTDFTDDEIKSAGNDKNTKTAFLDKLIKLDGSLKNVKAAKIISGDAETNKMLQ 103
      : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 99 -----EYCNFCGAYFVLENPNPRDIKFDDERPHKRRKRSKSSQSKSQTNRNRSQSNANAH 154
      : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 104 MLGTNATSFNSRNGTGEKKKKKKVKKEDKDEEKSTTKKRSKKEKHEKEKSEKKKS 163
      : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 155 FTSK-RRKDTKRRQERHIKEEDKEMTSK 183
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 164 AEEKKKKKSSSKRHKHKSDDRSSEKSEK 193
      | : : : : | : : : : | : : : : | : : : : | : : : : |

RESULT 19
JC5314
CDC28/cdc2-like kinase associating arginine-serine cyclophilin - human
N:Alternate names: CAR5-Cyp
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: JC5314; G02262
R:Nestel, F.P.; Colwill, K.; Harper, S.; Pawson, T.; Anderson, S.K.
Gene 180, 151-155, 1996
A:Title: RS cyclophilins: Identification of an NK-TR1-related cyclophilin.
A:Reference number: JC5314; MUID:97128820; PMID:8973360
A:Accession: JC5314
A:Molecule type: mRNA
A:Residues: 1-754 <NES>
A:Cross-references: EMBL:U40763; NID:g1117967; PIDN:AAB40347.1; PID:g1117968
A:Experimental source: thymus
A:Note: submitted to the EMBL Data Library, November 1995
C:Comment: This protein contains a cyclophilin-related domain, two 140K nuclear phospho
      splicing by binding to splicing factors containing serine-arginine repeats protein.
C:Genetics:
A:Gene: GDB:CYP; CAR5-CYP
A:Cross-references: GDB:9956062
C:Superfamily: CAR5 cyclophilin; cyclophilin homology
F:7-177/Domain: cyclophilin homology <CYP>

Query Match 9.8%; Score 99; DB 1; Length 754;
Best Local Similarity 21.7%; Pred. No. 2.7;
Matches 41; Conservative 25; Mismatches 51; Indels 72; Gaps 7;

QY 61 DFIV--GDW-----GNEQLRLRGFYKDASTIRKNSRISRLDYIKEYCN-----FGCAY 107
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 71 DFMVGGDFSEGNGRGGESI-YGFFFEDESFAVKHNK-----EFLLSMANRGRKDTNGSQF 124
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 108 FVLENPNPR-----D-----DIK 119
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 125 FITTKPTPHLDGHHVFGQVISQGVVREIENQKTDAAKPPFAEVRILSGELIPKSKVK 184
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 120 FDDERPHKRRKRSKSKS-----QSSKSQTRNNRSQSNANAHFTSKRKRKDTKRRQERHIKEE 174
      : : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 185 KEKKRKHKSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 244
      : : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 175 QDKEMTSK 183
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 245 KKKRKKSKK 253
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 20
A40459
nuclear phosphoprotein SRP55 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 24-Sep-1999
C:Accession: A40459; S14620
R:Roth, M.B.; Zahler, A.M.; Stolk, J.A.
J. Cell Biol. 115, 587-596, 1991
A:Title: A conserved family of nuclear phosphoproteins localized to sites of polymera
      A:Reference number: A40459; MUID:92011900; PMID:1717489
A:Accession: A40459
A:Status: preliminary
```

Qy 106 AYFVLE----NPNPRDIKFDDERPHKRRKSRKSQS-----SKSQTRNNRSQSNANAHF 155



AH1744  
B: subtilis yutD protein homolog lin2501 [imported] - Listeria innocua (strain Clip11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 17-May-2002  
C:Accession: AH1744  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AH1744  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC97728.1; PID:g16415023; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin2501  
C:Superfamily: Bacillus subtilis hypothetical protein yutD  
Query Match 23.7%; Score 239.5; DB 2; Length 91;  
Best Local Similarity 51.2%; Pred. No. 1.5e-12;  
Matches 43; Conservative 22; Mismatches 14; Indels 5; Gaps 2;  
Qy 31 DIEFQVLVINEKSAFDVTVFGQRFSEILLKYDFIVGDWGNQQLRLRGFYKDASTIRK---N 87  
Db 7 DLNVEITNYDAFDEEKLNRFSIDILGRYDVIWGDWGYDQLRLKGFEDDN--RKAAYD 64  
Qy 88 SRISLEDYIKVYCNFGCAFFVLE 111  
Db 65 NKISTLKDEIYVYCNFGCAFFVIR 88  
RESULT 7  
A:Accession: AB4079  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and Nucleic Acids Res. 28, 4317-4331, 2000  
A:Reference number: AB3650; MUID:20512582; PMID:11058132  
A:Accession: AB4079  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-93 <STO>  
A:Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BA07152.1; GSPDB:GN00178  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH3433  
C:Superfamily: Bacillus subtilis hypothetical protein yutD  
Query Match 23.2%; Score 234.5; DB 2; Length 93;  
Best Local Similarity 49.4%; Pred. No. 3.9e-12;  
Matches 44; Conservative 21; Mismatches 19; Indels 5; Gaps 2;  
Qy 26 IVKSDIEFQVLVINEKSAFDVTVFGQRFSEILLKYDFIVGDWGNQQLRLRGFYKDA---AS 82  
Db 1 MVRISNQFVIENVKDGNNEAFVRYSDVLNKYDVIWGDWGYDQLRLKGFEDDNKNS 60  
Qy 83 TIRKNSRISLEDYIKVYCNFGCAFFVLE 111  
Db 61 TYDK--KISTLDYIYVYCNFGCAFFVVK 87  
RESULT 8  
A:Accession: F70023  
A:Title: Complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: F70023  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber, C.; Bron, S.; Brouillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.E.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal, Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino, A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau, Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadiq, Y.; Sato, T.; Scani, A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, K.; Tosato, V.; Uchiya, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: F70023  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-112 <RUN>  
A:Cross-references: GB:299120; GB:AL009126; NID:g2635613; PIDN:CAB15221.1; PID:g26357  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yutD  
C:Superfamily: Bacillus subtilis hypothetical protein yutD  
Query Match 23.0%; Score 232.5; DB 2; Length 102;  
Best Local Similarity 51.7%; Pred. No. 6.2e-12;  
Matches 46; Conservative 16; Mismatches 26; Indels 1; Gaps 1;  
Qy 24 ENIVK:DDIEFQVLVINEKSAFDVTVFGQRFSEILLKYDFIVGDWGNQQLRLRGFYKDA- 82  
Db 10 EIMIL:QNAEEFVHNFKDGFNEEAFKARYSDILNKYDVIWGDWGYDQLRLKGFDDQNG 69  
Qy 83 TIRKNSRISLEDYIKVYCNFGCAFFVLE 111  
Db 70 KATFE:KISTLDYIYVYCNFGCAFFVLE 98  
RESULT 9  
A:Accession: S64146  
A:Title: probable membrane protein YGL133w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein G2842  
C:Species: Saccharomyces cerevisiae  
C:Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 19-Apr-2002  
C:Accession: S64146; S71739  
R:Escaribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64144  
A:Accession: S64146  
A:Molecule type: DNA  
A:Residues: 1-1264 <ESC>  
A:Cross-references: EMBL:Z72655; NID:gl322701; PID:e243361; PID:gl322702; MIPS:YGL133  
A:Experimental source: strain S288C  
R:Escaribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.  
Yeast 12, 887-892, 1996  
A:Title: Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces cerevisiae chro  
A:Reference number: S71733; MUID:96437978; PMID:8840506  
A:Accession: S71739  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1264 <ESW>  
A:Cross-references: EMBL:X92670  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995  
C:Genetics:  
A:Gene: SGD:ITC1  
A:Cross-references: SGD:S0003101  
A:Map position: 7L  
C:Keywords: transmembrane protein  
F:549-565/Domain: transmembrane #status predicted <TM>



## RESULT 6

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:51:32 ; Search time 17 Seconds  
(without alignments)  
1085.753 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 1009

Sequence: 1 MRKEVTPMLNKNYPGPQF.....EEQDKEMTSKQHLLEVRKN 192

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	452.5	44.8	176	C97956	hypothetical prote
2	436.5	43.3	170	H95088	conserved hypotet
3	331.5	32.9	251	B86647	hypothetical prote
4	241.5	23.9	127	F98558	conserved hypotet
5	239.5	23.7	91	A81375	B. subtilis YutD p
6	239.5	23.7	91	A81744	B. subtilis YutD p
7	234.5	23.2	93	A84079	hypothetical prote
8	232.5	23.0	102	F70023	hypothetical prote
9	109.5	10.9	1264	S84146	probable membrane
10	105	10.4	553	T27245	hypothetical prote
11	105	10.4	659	T27246	hypothetical prote
12	101	10.0	508	E71620	hypothetical prote
13	100.5	10.0	568	H88904	protein Y5G11C.9
14	100	9.9	269	B47112	growth response pr
15	99.5	9.9	272	S59042	splicing factor SR
16	99	9.8	484	A40988	54K arginine-rich
17	99	9.8	494	A48133	pre-mRNA splicing
18	99	9.8	535	T37189	hypothetical prote
19	99	9.8	754	JC5314	CDC28/cdc2-like ki
20	95	9.4	350	A40459	nuclear phosphop
21	95	9.4	712	T25438	hypothetical prote
22	94.5	9.4	755	S58718	probable nuclear p
23	94.5	9.4	1650	T18444	hypothetical prote
24	93	9.2	1294	T18473	hypothetical prote
25	93	9.2	1507	B47328	natural killer cel
26	92	9.1	344	S59043	splicing factor SR
27	91.5	9.1	451	E70013	hypothetical prote
28	91.5	9.1	1403	A47328	natural killer cel
29	90.5	9.0	707	S57157	hypothetical prote

30	90.5	9.0	1203	2	S27545	pullulanase - Ther
31	90	8.9	462	2	C84473	probable protein k
32	90	8.9	470	2	T34402	hypothetical prote
33	90	8.9	2573	2	D71614	hypothetical prote
34	89.5	8.9	374	2	A37282	52K active chroma
35	89.5	8.9	1827	2	T16270	hypothetical prote
36	89	8.8	443	2	C88427	protein R07E5.6 [i
37	89	8.8	592	2	S43597	coiled-coil protei
38	89	8.8	686	1	A44842	CGMP-gated ion cha
39	89	8.8	846	2	JC7720	acetyltransferase
40	89	8.8	1788	2	T31095	vitellogenin precu
41	89	8.8	1997	2	F71607	DNA helicase II BR
42	88.5	8.8	382	1	VHIHPC	nucleocapsid prote
43	88.5	8.8	690	2	A42161	CGMP-gated cation
44	88	8.7	490	2	T43745	clt4 protein - fls
45	87.5	8.7	299	2	T25596	hypothetical prote

ALIGNMENTS

RESULT 1

C97956

hypothetical protein spr0675 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C:Accession: C97956

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: C97956

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-176 <CUR>

A:Cross-references: GB:AE007317; PIDN:AAK99479.1; PID:g15458263; GSPDB:GN00174

C:Genetics:

A:Gene: spr0675

Query Match 44.8%; Score 452.5; DB 2; Length 176;  
Best Local Similarity 47.6%; Pred. No. 2.7e-29;  
Matches 91; Conservative 35; Mismatches 50; Indels 15; Gaps 4;

Qy 1 MRKEVTPMLNKNYPGPQFIHFENIVKSDDETFOLVINEKSAFDVTVFGQRFSEILKY 60

Db 1 MRKETAPLYNKNYPGPGEFHLGDKVETEGIAFSLVENIKDAFDVTFNQRFSEVLTKF 60

Qy 61 DFIVGNGNEOLRLRGFYKDASTIRKNSRISLEDIYKEYCNFGCAYFVLENPNDIKF 120

Db 61 DFIVGNGNEQLRLRGFYKDDRTKLEKISRLQDYLLEYCSYGCAFYVLENEAPRRASF 120

Qy 121 DDEREHKRRKRSKSSQSSKTSOTNRNRSQSNANAHFTSKRKDKTKRQERHIEEQDKENT 180

Db 121 D--KFWRTTEETPSRKGGKKTQTKR-KSNAD-----KKRRRQKDQHSQKEDKE-- 167

Qy 181 SAKQELLFVRK 191

Db 168 --QRFVIRQK 176

RESULT 2

H95088

conserved hypothetical protein sp0767 [imported] - Streptococcus pneumoniae (strain 1

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: H95088

R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp  
nson, T.; Hickley, E.K.; Holt, I.E.

Science 293, 448-506, 2001

**THIS PAGE BLANK (USPTO)**



CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF023142; AAD09327.1; -;  
 DR EMBL; AP001711; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AL117417; CAB55911.1; -;  
 DR EMBL; AB032998; BAA86486.1; -;  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 DR PROSITE; PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 KW RNA-binding; Alternative splicing.  
 FT NON\_TER 1 1  
 FT DOMAIN 5 8 POLY-ALA.  
 FT DOMAIN 165 168 POLY-PRO.  
 FT DOMAIN 314 321 POLY-ALA.  
 FT DOMAIN 323 326 POLY-PRO.  
 FT DOMAIN 518 592 RNA-BINDING (RRM).  
 FT DOMAIN 726 733 POLY-PRO.  
 FT DOMAIN 750 756 POLY-PRO.  
 FT DOMAIN 967 976 POLY-GLN.  
 FT VARSPLIC 776 798  
 FT  
 FT CONFLICT 270 274 MISSING (IN REF. 3).  
 FT CONFLICT 293 293 T -> I (IN REF. 1).  
 FT CONFLICT 296 296 T -> A (IN REF. 1).  
 FT CONFLICT 309 309 T -> A (IN REF. 1).  
 FT CONFLICT 312 312 V -> L (IN REF. 4).  
 FT CONFLICT 562 562 H -> Y (IN REF. 1).  
 FT CONFLICT 570 570 L -> V (IN REF. 1).  
 FT CONFLICT 581 581 N -> T (IN REF. 2).  
 FT CONFLICT 608 610 GWT -> CVI (IN REF. 1).  
 FT CONFLICT 614 616 WDK -> CON (IN REF. 1).  
 FT CONFLICT 622 622 L -> M (IN REF. 1).  
 FT CONFLICT 626 626 C -> V (IN REF. 1).  
 FT CONFLICT 818 836 AVPPAAPTLNPPPPVQPV -> TCCTHSAHPSCNPACLP  
 C (IN REF. 1).  
 FT CONFLICT 837 866 SLGTCGVAPGPVIGLQEPSTGLGAPGL -> VSLLGKL  
 FFIIIFPCPTPPPHPTHPQVRF (IN REF. 3).  
 FT CONFLICT 854 854 A -> E (IN REF. 1).  
 FT CONFLICT 1149 1149 S -> G (IN REF. 1).  
 SQ SEQUENCE 1157 AA; 126846 MW; 176A7789998BF01B CRC64;  
 Query Match 8.4%; Score 84.5; DB 1; Length 1157;  
 Best Local Similarity 33.3%; Pred. No. 22;  
 Matches 20; Conservative 14; Mismatches 25; Indels 1; Gaps 1;  
 QY 124 RPHKRRKRSKSSQTSQTRNNSSNANAHFTSKRKDKTKRQERHIKEEQDKEMTSK 183  
 Db 451 RSPARRRSKRSRRSRHRRSRSSRRDRRHSPRSQRRDRKE-RRRQKGLPQVK 509  
 RESULT 40  
 NCR2\_MOUSE  
 ID NCR2\_MOUSE STANDARD; PRT; 2472 AA.  
 AC Q9WU42; Q9WU43; Q9WUC1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nuclear receptor co-repressor 2 (N-COR2) (Silencing mediator of  
 DE retinoic acid and thyroid hormone receptor) (SMRT) (SMRTE) (Thyroid-,  
 DE retinoic-acid-receptor-associated co-repressor) (TRAC).  
 DE associating factor) (TRAC).  
 GN NCR2 OR SMRT.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).  
 RC TISSUE=Splice, and Brain;  
 RX MEDLINE=99178941; PubMed=10077563;  
 RA Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;  
 RT "Unique forms of human and mouse nuclear receptor corepressor SMRT.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA).  
 RC TISSUE=Embryo;  
 RX MEDLINE=99199215; PubMed=10097068;  
 RA Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;  
 RT "SMRte, a silencing mediator for retinoid and thyroid hormone  
 RT receptors-extended isoform that is more related to the nuclear  
 RT receptor corepressor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).  
 CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
 CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
 CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
 CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIIB.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS. ALSO WIDELY EXPRESSED IN EARLY  
 CC EMBRYOS.  
 CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
 CC DOMAINS (ID1 AND ID2).  
 CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND  
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES  
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
 CC SPECIFICITY.  
 CC -1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CORNR BOXES.  
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF113001; AAD20944.1; -;  
 DR EMBL; AF113002; AAD20945.1; -;  
 DR EMBL; AF125671; AAD22972.1; -;  
 DR TRANSFAC; T04690; -;  
 DR MGD; MGI:1337080; Ncor2.  
 DR InterPro; IPR001005; Myb\_DNA\_binding.  
 DR Pfam; PF00249; myb\_DNA-binding; 2.  
 DR SMART; SM00395; SANT; 2.  
 DR PROSITE; PS50090; MYB\_3; 1.  
 KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 KW Coiled coil; Alternative splicing.  
 FT DOMAIN 165 207 COILED COIL (POTENTIAL).  
 FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).  
 FT DNA\_BIND 429 474 SANT-A (POTENTIAL).  
 FT MYB 609 653 MYB.  
 FT COILED COIL (POTENTIAL).  
 FT COILED COIL (POTENTIAL).  
 FT COILED COIL (POTENTIAL).  
 FT PRO-RICH. 775 804

GN Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND FUNCTION.  
 RX MEDLINE=87301717; PubMed=2441872;  
 RA Boggs R.T., Gregor P., Idriss S., Belote J.M., McKeown M.;  
 RT "Regulation of sexual differentiation in *D. melanogaster* via  
 alternative splicing of RNA from the transformer gene.";  
 RL Cell 50:739-747(1987).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94283902; PubMed=8013913;  
 RA Walther C.S., Schaeffer S.W.;  
 RT "Molecular population genetics of sex determination genes: the  
 transformer gene of *Drosophila melanogaster*.";  
 RL Genetics 136:1367-1372(1994).  
 RN (3)  
 RP CHARACTERIZATION OF RS DOMAIN.  
 RX MEDLINE=92005712; PubMed=1655279;  
 RA Li H., Bingham P.M.;  
 RT "Arginine/serine-rich domains of the su(wa) and tra RNA processing  
 regulators target proteins to a subnuclear compartment implicated in  
 splicing.";  
 RL Cell 67:335-342(1991).  
 RN (4)  
 RP INTERACTION WITH SR PROTEINS IN ENHANCER COMPLEX.  
 RX MEDLINE=93327418; PubMed=8334698;  
 RA Tian M., Maniatis T.;  
 RT "A splicing enhancer complex controls alternative splicing of  
 doublesex pre-mRNA";  
 RL Cell 74:105-114(1993).  
 CC -!- FUNCTION: Member of the regulatory pathway controlling female  
 somatic sexual differentiation, regulated by Sxl. Activates  
 dsx female-specific splicing by promoting the formation of a  
 splicing enhancer complex which consists of tra, tra2 and sr  
 proteins.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR; SPECKLED SUBNUCLEAR COMPARTMENT.  
 CC -!- DOMAIN: RS DOMAIN DIRECTS LOCALIZATION OF PROTEINS TO THE SPECKLED  
 SUBNUCLEAR COMPARTMENT AND THE PURPOSE OF THIS LOCALIZATION IS TO  
 ALLOW COLOCALIZATION AND CO-CONCENTRATION OF COMPONENTS OF THE  
 SPLICING AND SPLICING REGULATORY MACHINERY TO PERMIT RELATIVELY  
 HIGH RATES AND/OR EFFICIENCIES OF REACTION AND INTERACTION.  
 CC -!- MISCELLANEOUS: The sexual regulation of tra occurs through a  
 mechanism of sex-specific alternative RNA splicing. The non-sex-  
 specific RNA expressed in males is not translated.  
 CC -!- CAUTION: It is uncertain whether Met-1 or Met-3 is the initiator.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed, usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L19465; AAA75341.1; -;  
 DR EMBL; L19466; AAA75342.1; -;  
 DR EMBL; L19467; AAA28958.1; -;  
 DR EMBL; L19468; AAA28959.1; -;  
 DR EMBL; L19469; AAA28960.1; -;  
 DR EMBL; L19470; AAA28961.1; -;  
 DR EMBL; L19618; AAA28962.1; -;  
 DR EMBL; L19619; AAA28963.1; -;  
 DR EMBL; L19620; AAA28964.1; -;  
 DR EMBL; M17478; AAB59226.1; -;  
 DR EMBL; L19464; AAA75340.1; -;  
 DR PIR; A29648;  
 DR Flybase; FBgn0003741; tra.  
 KW Nuclear protein; Sexual differentiation.

FT DOMAIN : 15 125 ARG/SER-RICH (RS DOMAIN).  
 SQ SEQUENCE 197 AA; 24134 MW; 10C0B7A6F0C1AF81 CRC64;  
 Query Match 8.4%; Score 84.5; DB 1; Length 197;  
 Best Local Similarity 20.4%; Pred. No. 3.1;  
 Matches 22; Conservative 20; Mismatches 35; Indels 31; Gaps 2;  
 Qy 81 ASTIRKNSRISRLDYIKCYCNFCAYFVLENPNPNRDIKFDDERPH---KRRKSRKSKSQS 137  
 Db 48 ADEVREQRLRLRLR-----QRAHQSTRTRTSRSRSQS 79  
 Qy 138 SKSTQNNRNSQSNANAHFTSKKRDKTKRQERHIKEQDKWTSKQK 185  
 Db 80 SIRESPFHRHRQSRNRNRSRNRSSRKRQRORSRSRSSRRRRQRSPH 127  
 RESULT 39  
 SRAM\_HUMAN  
 ID SRA4\_HUMAN STANDARD; PRT; 1157 AA.  
 AC O95104; O95F01; Q9ULP8;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE CTD-binding SR-like protein RA4 (Fragment).  
 GN KIAA1172.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Tassone F., Villard L., Gardiner K.;  
 RT "Sequence, genomic organization and map localization of the human SR  
 protein gene ra4.";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20389799; PubMed=10830953;  
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
 RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,  
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,  
 RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,  
 RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,  
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,  
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
 RA Leirach H., Reinhardt R., Yaspo M.-L.;  
 RT "The DNA sequence of human chromosome 21.";  
 RL Nature 405:311-319(2000).  
 RN (3)  
 RP SEQUENCE OF 195-866 FROM N.A.  
 RC TISSUE=Testis;  
 RA Ottenwaelder B., Obermaier B., Mewes H.-W., Gassenhuber J.,  
 RA Wiemann S.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP SEQUENCE OF 207-1157 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20039618; PubMed=10574461;  
 RA Hirosewa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,  
 RA Ohara O.;  
 RT "Characterization of cDNA clones selected by the Genemark analysis  
 from size-fractionated cDNA libraries from human brain.";  
 RL DNA Res. 6:329-336(1999).  
 CC -!- FUNCTION: MAY ACT TO PHYSICALLY AND FUNCTIONALLY LINK  
 TRANSCRIPTION AND PRE-MRNA PROCESSING (BY SIMILARITY).  
 CC -!- SUBUNIT: INTERACT WITH THE REPETITIVE C-TERMINAL DOMAIN (CTD) OF  
 RNA POLYMERASE II (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE  
 PROBABLY PRODUCED BY ALTERNATIVE SPLICING.







```

QY 173 EQDKEMTSKAK 184
ID YF13_YEAST STANDARD; PRT: 787 AA.
DB 689 EERGGKSKSRK 700

RESULT 33
YF13_YEAST
AC P43596;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 90.9 kba protein in GCN20-CMK1 intergenic region.
GN YF013W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae."
RL Nat. Genet. 10:261-268(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50617; BAA09252.1; -
DR SGD; S0001909; IOC3.
KW Hypothetical protein.
SQ SEQUENCE 787 AA; 90896 MW; 1919A8A1F58B5340 CRC64;

Query Match 8.6%; Score 86.5; DB 1; Length 787;
Best Local Similarity 24.1%; Pred. No. 9.8;
Matches 27; Conservative 22; Mismatches 42; Indels 21; Gaps 4;

QY 69 NEQLRLRGFYKDA--STIRKNSRISRLDYIKCYCNFGCAYFVLENPNDRIKFDDEPH 126
DB 682 NFSINTNPDKAKYATARKKIQI--LKELFSDY-----YFILFOEQMKVOFADMKPG 733
QY 127 KRRKRSKSSQSSKQTRNNRSQSNANAHFTSKRRKDTKRRQERHIKEEQDKE 178
DB 734 KROLRIQRQT-----VNYTEYDSEYVDDEEDDEADIYDDNDND 774

RESULT 34
DDX8_HUMAN
ID DDX8_HUMAN STANDARD; PRT: 1220 AA.
AC Q14562;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent helicase DDX8 (RNA helicase HRH1) (DEAH-box protein 8).
GN DDX8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021291; PubMed=7935475;
RA Ono Y., Ohno M., Shimura Y.;

"Identification of a putative RNA helicase (HRH1), a human homolog of
yeast Prp22."
Mol. Cell. Biol. 14:7611-7620(1994).
[2]
RP FUNCTION.
RX MEDLINE=96196415; PubMed=8608946;
RA Ohno M., Shimura Y.;
RT "A human RNA helicase-like protein, HRH1, facilitates nuclear export
RT of spliced mRNA by releasing the RNA from the spliceosome."
Genes Dev. 10:997-1007(1996).
CC -!- FUNCTION: FACILITATES NUCLEAR EXPORT OF SPLICED MRNA BY RELEASING
CC THE RNA FROM THE SPLICESOME.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DOMAIN: THE RS DOMAIN CONFERS A NUCLEAR LOCALIZATION SIGNAL, AND
CC APPEARS TO FACILITATE THE INTERACTION WITH THE SPLICESOME.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAD
CC SUBFAMILY DDX8/PRP22 ORTHOLOG.
CC -!- SIMILARITY: CONTAINS 1 SL MOTIF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50487; BAA09078.1; -
DR HSP; P05055; LSRO.
DR Genew; HGNC:2749; DDX8.
DR MIM; 600396; -
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAH_box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR003029; SL.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00575; SL; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR SMART; SM00316; SL; 1.
DR PROSITE; PS00690; DEAH-ATP_HELICASE; 1.
DR PROSITE; PS0126; SL; 1.
KW mRNA processing; mRNA splicing; Helicase; ATP-binding;
KW Nuclear protein.
FT DOMAIN 172 175 POLY-LYS.
FT DOMAIN 176 228 ARG/SER-RICH (RS DOMAIN).
FT DOMAIN 265 336 SL MOTIF.
FT NP_BIND 588 595 ATP (POTENTIAL).
FT SITE 685 688 DEAH BOX.
FT MUTAGEN 594 594 K->E: IN GET; INHIBITION OF PRE-MRNA
FT SPlicing AND NUCLEAR EXPORT OF UNSPLICED
FT RNA.
FT MUTAGEN 717 717 S->L: IN LAT; INHIBITION OF PRE-MRNA
FT SPlicing AND NUCLEAR EXPORT OF UNSPLICED
FT RNA.
SQ SEQUENCE 1220 AA; 139314 MW; 17C1602A73A0EF24 CRC64;

Query Match 8.6%; Score 86.5; DB 1; Length 1220;
Best Local Similarity 35.3%; Pred. No. 16;
Matches 24; Conservative 18; Mismatches 21; Indels 5; Gaps 3;

QY 116 RDIKFDDERPHKRR-KSRKSSQSSKQTRN---NRSQSNANAHFTSKRRKDTKRRQERHI 171
DB 188 RDRERNRDRDHKRRHRSR-SRTREKNVKRSYRSRSRSQSPPKDRKDRDKYGERNL 246
QY 172 KEEDQKEM 179
DB 247 DRWRDKHV 254

RESULT 35
COTB_BACSU
ID COTB_BACSU STANDARD; PRT: 380 AA.

```

```

ID CYLL_HUMAN STANDARD; PRT; 598 AA.
AC P35663;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE Cyclicin I (Multiple-band polypeptide I) (Fragment).
GN CYLC1 OR CYLI OR CYL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Br15:01 N2;
RX MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
RT "Molecular characterization of mammalian cyclicin, a basic protein of
the sperm head cytoskeleton."
RL J. Cell Biol. 122:1043-1052(1993).
CC -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CYLIX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -!- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22780; CAA80457.1; -
DR PIR; S35920; S35920.
DR PIR; B40713; B40713.
DR Genew; HGNC:2582; CYLC1.
DR MIM; 603121; -
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT NON_TER 1 1
FT DOMAIN 479 500 8 APPROXIMATE TANDEM REPEATS.
FT REPEAT 225 252 1.
FT REPEAT 253 289 2.
FT REPEAT 290 326 3.
FT REPEAT 327 364 4.
FT REPEAT 365 400 5.
FT REPEAT 401 438 6.
FT REPEAT 439 478 7.
FT REPEAT 479 500 8.
FT DOMAIN 548 598 PRO-RICH.
SQ SEQUENCE 598 AA; 68034 MW; 062BA2E2D2AB61F7 CRC64;

Query Match 8.6%; Score 86.5; DB 1; Length 598;
Best Local Similarity 21.6%; Pred. No. 7.3;
Matches 38; Conservative 38; Mismatches 71; Indels 29; Gaps 7;

QY 13 NKYPGPQFIHFENIVKSDIEFQVINEKSAFDVTFVQGRFSEILLKYDF--IVGDWGN 70
DB 153 SKTNTFELTKNNPKKD-----LKRKTSND--PISEICSENLSNVDFLMLVGQSDDE 204
QY 71 QLRLRGYKDASTRKRSRISRLDYKEYCNF-----GCAYFVLEPNPNRDIKFD 122
DB 205 SINFDALRWYS--QNSK-----NYSKYTKYTKDKTKNAKSSDAESDSDAKKDS 257
QY 123 ERPHKRKRKSSQSSQTSNNRQSNANAHFTSKRKDKTKRQERHIEEQDK 178
DB 258 KVKVKNVKKDDKKDKKDKTESTDAESG----SKDERKDKTKDKKKLKKDKKKD 309

RESULT 32
Y061_CAEEL
ID Y061_CAEEL STANDARD; PRT; 724 AA.
AC P34600;

DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical. 84.7 kDa protein ZK1098.1 in chromosome III.
GN ZK1098.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Br15:01 N2;
RX MEDLINE=94130718; PubMed=7906398;
RA Willson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: CONTAINS 2 WW DOMAINS.
CC -!- SIMILARITY: TO S.POMBE SPAC4D7.13.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22176; CAA80142.1; -
DR PIR; S40923; S40923.
DR WormPep; ZM1098.1; CE03847.
DR InterPro; IPR002713; FF.
DR Pfam; PF00397; WW; 2.
DR Pfam; PF01446; FF; 5.
DR SMART; SM00441; FF; 3.
DR SMART; SM00456; WW; 2.
DR PROSITE; P01159; WW_DOMAIN_1; 2.
DR PROSITE; P50020; WW_DOMAIN_2; 2.
KW Hypothetical protein; Repeat.
FT DOMAIN 78 111 WW 1.
FT DOMAIN 123 156 WW 2.
SQ SEQUENCE 724 AA; 84664 MW; BCC7AACAD5AA4C20 CRC64;

Query Match 8.6%; Score 86.5; DB 1; Length 724;
Best Local Similarity 17.5%; Pred. No. 9;
Matches 44; Conservative 45; Mismatches 84; Indels 79; Gaps 8;

QY 2 RKEVTPEMLNKNYKPGQFIHFENIVKSDIEFQVINEKSAFDVTFV----- 49
DB 459 RGLTMSLWTSLFP-----IISTDTRFELMLFQPGSSPLDLFKFFVEDLKEQYT 508
QY 50 --GQTSSEILLK-----YDFIVG-----DWGNSQLRLRGFYKDA----- 81
DB 509 EDRRLKEILTEKGCQVIATTEYREFSDVWVSHKGGKVDHGNMKLCYNLSLIEAKSKAK 568
QY 82 ----STIRKNSRI--SRLEDYIKY-----CNFGCAYFVLEPNPNRDIKFD 122
DB 569 DEEKEJLRKRRLSEFEFNLLKEHNVKDSWTVIKPIEKDKAYLAMENDERETAFNH 626
QY 123 ERPHK-----RRKRSKSSQSSQTSNNRQSNANAHFTSKRKDKTKRQERH 172
DB 629 YKNGTSGTTAGSETLEKRRKKKKKKKRNKSDNNSESEGEIREKREKKKKKKHSDRMD 688
```





	-!	SUBCELLULAR LOCATION:	Type II membrane protein (Probable).
CC	-!	SIMILARITY:	TO YEAST YJR061W.
CC	-!	CATION:	REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
CC			-----
CC	This SWISS-PROT entry is copyright.	It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isib-sib.ch/announce/or send an email to licensel@isib-sib.ch).	-----
CC	EMBL:	D83006;	BAA11676.1; --
DR	EMBL:	Z28201;	CAAG82046.1; --
DR	EMBL:	Z28200;	CAA82044.1; --
PIR:	PIR:	S38037;	S38037.
DR	PIR:	S38038;	S38038.
DR	SGD:	S0001683;	MNN4.
KW	Transmembrane:	Signal-anchor; Repeat.	
FY	DOMAIN	1	27 CYTOPLASMIC (POTENTIAL), SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT	TRANSMEM	28	48
FF			
FT	DOMAIN	49	1178 LUMENAL (POTENTIAL).
FT	DOMAIN	1032	1174 ARG/GLU/LYS-RICH (HIGHLY CHARGED).
FT	DOMAIN	1042	1174 17 X 8 AA TANDEM REPEATS OF K-K-K-E-E-E-
FT	REPEAT	1042	1049 1.
FT	REPEAT	1050	1057 2.
FT	REPEAT	1058	1065 3.
FT	REPEAT	1066	1073 4.
FT	REPEAT	1074	1081 5.
FT	REPEAT	1082	1089 6.
FT	REPEAT	1090	1097 7 (APPROXIMATE).
FT	REPEAT	1098	1105 8.
FT	REPEAT	1106	1113 9 (APPROXIMATE).
FT	REPEAT	1114	1121 10 (APPROXIMATE).
FT	REPEAT	1122	1129 11 (APPROXIMATE).
FT	REPEAT	1130	1137 12.
FT	REPEAT	1138	1144 13 (APPROXIMATE).
FT	REPEAT	1145	1152 14 (APPROXIMATE).
FT	REPEAT	1153	1160 15 (APPROXIMATE).
FT	REPEAT	1161	1168 16 (APPROXIMATE).
FT	REPEAT	1169	1174 17 (INCOMPLETE).
FT	DOMAIN	37	40 POLY-IIE.
FT	DOMAIN	1042	1045 POLY-LYS.
FT	DOMAIN	1046	1049 POLY-GLU.
FT	DOMAIN	1050	1053 POLY-LYS.
FT	DOMAIN	1054	1057 POLY-GLU.
FT	DOMAIN	1058	1061 POLY-LYS.
FT	DOMAIN	1062	1065 POLY-GLU.
FT	DOMAIN	1066	1069 POLY-LYS.
FT	DOMAIN	1070	1073 POLY-GLU.
FT	DOMAIN	1074	1077 POLY-LYS.
FT	DOMAIN	1078	1081 POLY-GLU.
FT	DOMAIN	1082	1085 POLY-LYS.
FT	DOMAIN	1086	1089 POLY-GLU.
FT	DOMAIN	1094	1097 POLY-GLU.
FT	DOMAIN	1098	1101 POLY-LYS.
FT	DOMAIN	1102	1105 POLY-GLU.
FT	DOMAIN	1134	1137 POLY-GLU.
FT	DOMAIN	1157	1160 POLY-GLU.
FT	DOMAIN	1165	1168 POLY-GLU.
SQ	SEQUENCE	1178 AA;	I39380 MW; BC05DAE0AEFCB282 CRC64;
	Query Match	8.6%;	Score 87; DB 1; Length .1178;
	Best Local Similarity	18.7%;	Pred. No. 14;
	Matches	34; Conservative	34; Mismatches 76; Indels 38; Gaps 5;
OQ	19 QTFHFENVKSDIDFOLVINEKSADFVFVGQRSEILLKYDFIVGDWGNEQLRLRGFY 78 :   :   I :	 I :   I :   I :   I :	----
Db	924 KVIVEEVGVIIDLSELSKYRNLSFNIT	----FDETGTTLDDNTD	---y 967

```
DR Pfam: PF00642; zf-COCH; 2.
DR SMART: SM00360; RRM; 1.
DR SMART: SM00356; ZnF_C3H1; 2.
DR PROSITE: PS0102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP1; FALSE_NEG.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Zinc-finger; Repeat.
FT DOMAIN 189 295 RNA-BINDING (RRM).
FT DOMAIN 85 88 POLY-LYS.
FT DOMAIN 116 119 POLY-GLU.
FT CONFLICT 351 351 N -> S (IN REF. 2).
SQ SEQUENCE 428 AA; 51364 MW; 4542EE234B0759E6 CRC64;

Query Match
Best Local Similarity 8.7%; Score 87.5; DB 1; Length 428;
Matches 31; Conservative 9; Mismatches 44; Indels 31; Gaps 5;

QY 99 EYCNFGCAVYVLENPN-----PRDI-----KFDDEPHKRR-----KRSK 134
Db :||| | ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
315 KHCNF---LHVFNPNNPEANRDIYMSPPAWTGGSGKNSDRERKDHHEEYYSKRSY 371
QY 135 SOSKSOTNRNRSOSANAHFTSKRKDKPKRQERHI-----KEODKEMTSAKOH 185
Db :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
372 HSGSYHSSARNRESERKSPHRWKKSHKQTKSHERHSSRRGREDDSSPGPQSOSH 426

RESULT 25
EX5B_BORBU
ID EX5B_BORBU STANDARD; PRT; 1169 AA.
AC 051578;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).
GN RECB OR B0633.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Cocayne J.D., Weidman J.,
RA Utterback T., Wathey L., McDonald L., Artlich P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
burgdorferi".
RL Nature 390:580-586(1997).
CC -!- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
(BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
phosphooligonucleotides.
CC -!- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
(BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
DR HSP; P56255; LPJR.
DR TIGR: BB0633; -
DR InterPro: IPR004586; RecB.
DR InterPro: JPR000212; UvrD-helicase.
DR Pfam: PF00580; UvrD-helicase; 1.
DR TIGRFAMS: TIGR00609; recB; 1.
KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
DNA repair; Complete proteome.
FT NP_BIND 18 25 ATP (POTENTIAL).
SQ SEQUENCE 1169 AA; 137828 MW; B61D63C1C959B91F CRC64;

Query Match
Best Local Similarity 8.7%; Score 87.5; DB 1; Length 1169;
Matches 48; Conservative 32; Mismatches 77; Indels 51; Gaps 9;

QY 8 EMLNTYIKY-PGQF-----THFENIVKSDIEFOLVINEKSAFDVTFP---GQRFSEIL 57
Db :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
109 ETENTY:KYKPKREKFEIDEIVDFLRKSDSLIQLDIKD---YELKVPKSDARKTEIV 165
QY 58 LKYDF-----VGDGNEQ-----LRLRGFYKDASTI-----RK 86
Db :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
166 LAIKKAYERTDQELGDMWKQTAFENILLKKEELIKDKIILEDLDMKTDKDEILSFYNK 225
QY 87 NSRISKLE-DYIKYCNFGCAVYVLENPNPRDIKEDDERPHKRRKRSKSSQSKSQTNRN 145
Db :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
226 HIQTGKLEIEYSKENDIFKIAETLLKN-----KFFSTLIEKETKNSKLSPELKIKND 279
QY 146 RSQSNINAHFTSKRKDKPKRQERHIKE 173
Db 280 LITGLINIKHEKYSKEDNRNKRNNLQ 307

RESULT 26
KMH_B_DICDI
ID KMH_B_DICDI, STANDARD; PRT; 732 AA.
AC P90648;
DT 15-JUL-1993 (Rel. 36, Created)
DT 15-JUL-1993 (Rel. 36, Last sequence update)
DT 15-JUL-1993 (Rel. 36, Last annotation update)
DE Myosin heavy chain kinase B (EC 2.7.1.129) (MHCK B).
GN MHCK OR MH2KB.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=97277316; PubMed=9115238;
RA Clancy C.E., Mendoza M.G., Naismith T.V., Kolman M.F., Egelhoff T.T.;
RT "Identification of a protein kinase from Dictyostelium with homology
to the novel catalytic domain of myosin heavy chain kinase A.";
RL J. Biol. Chem. 272:11812-11815(1997).
CC -!- FUNCTION: PHOSPHORYLATES THREONINE IN THE C-TERMINAL TAIL REGION
OF MYOSIN II HEAVY CHAIN. THIS PHOSPHORYLATION IS CRITICAL IN
REGULATING THE ASSEMBLY AND DISASSEMBLY OF MYOSIN II FILAMENT.
CC -!- CATALYTIC ACTIVITY: ATP + [myosin heavy-chain] = ADP + [myosin
heavy-chain] phosphate.
CC -!- DOMAIN: CONSISTS OF AN N-TERMINAL DOMAIN, A CENTRAL NONREPETITIVE
CATALYTIC DOMAIN, AND A C-TERMINAL DOMAIN WITH SEVEN WD REPEATS.
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE MHCK / EF-2 PROTEIN KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

CC COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN  
 CC BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO AN OPENING OF THE CATION  
 CC CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD  
 CC PHOTORECEPTORS.  
 CC  
 CC -1- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER. FORMS HETEROOLIGOMERIC  
 CC COMPLEX WITH CNG4.  
 CC  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC  
 CC -1- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.  
 CC  
 CC -1- DISEASE: DEFECTS IN CNGAI ARE A CAUSE OF AUTOSOMAL RECESSIVE  
 CC RETINITIS PIGMENTOSA (ARRP). A DISEASE THAT LEADS TO DEGENERATION  
 CC OF RETINAL PHOTORECEPTOR CELLS.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL  
 CC FAMILY.  
 CC  
 CC -1- DATABASE: NAME=Mutations of the CNGAI gene;  
 CC NOTE=Retina International's Scientific Newsletter;  
 CC WWW="http://www.retina-international.com/sci-news/cngalmut.htm".  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; M84741; AA52010.1; ALT\_INIT.  
 CC EMBL; S42457; AAB22778.1; -.  
 CC EMBL; S76062; AAD14206.1; -.  
 CC PIR; A42161; A42161.  
 CC Genew; HGNC:2148; CNGAI.  
 CC MTM; 123825; -.  
 CC InterPro; IPR000636; M+channel\_nlg.  
 CC InterPro; IPR000595; cNMP\_binding.  
 CC Pfam; PF00027; cNMP\_binding; 1.  
 CC Pfam; PF00520; ion\_trans; 1.  
 CC SMART; SM00100; cNMP; 1.  
 CC PROSITE; PS00888; cNMP\_BINDING\_1; 1.  
 CC PROSITE; PS00889; cNMP\_BINDING\_2; 1.  
 CC PROSITE; PS00042; cNMP\_BINDING\_3; 1.  
 CC Ionic channel; Ion transport; cAMP-binding; Transmembrane;  
 CC Multigene family; Vision; Disease mutation; Polymorphism;  
 CC Retinitis pigmentosa.  
 CC  
 CC FT DOMAIN 1 160 CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 161 181 H1 (POTENTIAL).  
 CC FT DOMAIN 182 194 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 195 213 H2 (POTENTIAL).  
 CC FT DOMAIN 214 237 CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 238 257 H3 (POTENTIAL).  
 CC FT DOMAIN 258 295 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 296 318 H4 (POTENTIAL).  
 CC FT DOMAIN 319 370 CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 371 390 H5 (POTENTIAL).  
 CC FT DOMAIN 391 474 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 475 495 H6 (POTENTIAL).  
 CC FT DOMAIN 496 685 CYTOPLASMIC (POTENTIAL).  
 CC FT NP\_BIND 483 605 CGMP (POTENTIAL).  
 CC FT BINDING 542 542 CGMP (POTENTIAL).  
 CC FT BINDING 557 557 CGMP (POTENTIAL).  
 CC FT CARBOHYD 421 421 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC FT VARIANT 28 28 R -> Q.  
 CC FT VARIANT 114 114 /FTId=VAR\_009295.  
 CC FT VARIANT 316 316 /FTId=VAR\_009296.  
 CC FT VARIANT 46 46 S -> F (IN ARR).  
 CC FT CONFLICT 85 85 L -> Y (IN REF. 1).  
 CC FT CONFLICT 146 147 EE -> HH (IN REF. 1).  
 CC FT CONFLICT 539 539 Y -> T (IN REF. 1).  
 CC FT CONFLICT 677 678 GA -> WS (IN REF. 1).  
 CC SEQUENCE 686 AA; 79126 MW; E520D216FC97AF6 CRC64;

Query Match

8.7%; Score 88; DB 1; Length 686;

Best Local Similarity 27.0%; Pred. No. 6.4;  
 Matches 31; Conservative 23; Mismatches 41; Indels 20; Gaps 6;  
 QY 69 NEQLRLRGFYKDASTRKNSRISLEDYIKYCNFGCAYFVLENPNDRIKFDDEPHKR 128  
 DB 53 NENPHARGSFYSYKS-LRKGGPSQR-----EQYLPGATALFNVNSSNK-----DOEPEEK 101  
 QY 129 RKRSKSSQSKSQTRNNRSQSNANAHFTSKRKDKTKRQE----RHKEEQDKEM 179  
 DB 102 KK-KKKEKSKSDKN---ENKNDPEKSKKKKKKKKEKSKDKKEEKKEV 151  
 RESULT 24  
 U2RL\_MOUSE  
 ID U2RL\_MOUSE STANDARD; PRT; 428 AA.  
 AC Q64707;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit  
 DE related-protein 1 (SP2).  
 GN U2AF1-RS1 OR SP2 OR SP2-7.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC [1]  
 RN NCBI\_TaxID=10090;  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RX MEDLINE=94111973; PubMed=8284201;  
 RA Hatada I., Sugama T., Mukai T.;  
 RT "A new imprinted gene cloned by a methylation-sensitive genome  
 RT scanning method.";  
 RL Nucleic Acids Res. 21:5577-5582(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PWK; TISSUE=Liver;  
 RX MEDLINE=94184361; PubMed=8136831;  
 RA Hayashizaki Y., Shibata H., Hirotsune S., Sugino H., Okazaki Y.,  
 RA Sasaki N., Hirose K., Imoto H., Okuzumi H., Muramatsu M.,  
 RA Komatsubara H., Shiroishi T., Moriaki K., Katsuki M., Hatano N.,  
 RA Sasaki H., Ueda T., Mise N., Takagi N., Plass C., Chapman V.M.;  
 RT "Identification of an imprinted U2af binding protein related sequence  
 RT on mouse chromosome 11 using the RLGS method.";  
 RL Nat. Genet. 6:33-40(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=95175367; PubMed=7870588;  
 RA Hatada I., Kitagawa K., Yamaoka T., Wang X., Arai Y., Hashido K.,  
 RA Ohishi S., Masuda J., Ogata J., Mukai T.;  
 RT "Allele-specific methylation and expression of an imprinted U2af1-rs1  
 RT (SP2) gene.";  
 RL Nucleic Acids Res. 23:36-41(1995).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -1- SIMILARITY: CONTAINS 2 C3H1-TYPE ZINC FINGERS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; D17407; BAA04230.1; -.  
 CC EMBL; S69507; AAB30301.1; -.  
 CC EMBL; D26474; BAA05486.1; -.  
 CC MGD; MGI:98885; U2af1-rs1.  
 CC InterPro; IPR000504; RNA\_rec\_mot.  
 CC InterPro; IPR000571; Znf\_CCCH.  
 CC Pfam; PF00076; rrm; 1.







```

DR InterPro; IPR000237; GRIP_domain.
DR Pfam; PF01465; GRIP; 1.
KW Golgi stack; Antigen; Coiled coil; Alternative splicing.
FT DOMAIN 133 237
FT DOMAIN 276 1011
FT DOMAIN 1033 1214
FT DOMAIN 1259 2152
FT VARSPLIC 2154 2185
FT TPYKGNLYHTDVSFLGPEFTEFYLRKVLPEY -> HLTKV
FT AICTIRMSHLENLPNLSICEKCFLSI (IN ISOFORM
FT 2).
FT MISSING (IN ISOFORM 2).
FT VARSPLIC 2186 2230
FT VARSPLIC 2103 2109
FT VARSPLIC 2222 2230
FT FTSPRSGLF -> SWLRSSS (IN ISOFORM 4).
FT CONFLICT 188 188
FT CONFLICT 220 220
FT CONFLICT 276 276
FT CONFLICT 584 584
FT CONFLICT 628 628
FT CONFLICT 630 630
FT CONFLICT 682 682
FT SEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;
Query Match 9.0%; Score 90.5; DB 1; Length 2230;
Best Local Similarity 20.1%; Pred. No. 15;
Matches 38; Conservative 47; Mismatches 75; Indels 29; Gaps 8;
Qy 12 YNKYPGPGFIHFENIVKSDDIIEFQVLVINEKSAFD-----VTFGQRFSEILLKYDFIVGDW 67
Db 1445 WKKAQSRFTQHTQTVK--ELQIOLEKSKAEYKDEQINLLKELDOONKRFDCCLKGEM 1502
Qy 68 GNEOLRLRGFYKDASTRKN--SRISRLDYI-KEYCNFGCAYFVLEPNP-PRDIKEDDE 123
Db 1503 EDDSKMEKKNESLTELKSTQARIMELEDHTQKTIIESLNEVLKYNQNOKDIE----- 1558
Qy 124 RPHKRRSRKSKQSKSOTRRNSQ-----SNANAHTSKRKDTRKRQERHI----- 171
Db 1559 --HKVLQKLFQFQELGEEKDNVRKAEKILITLENQVYSMAKAELETKKKELEHVNLSVK 1616
Qy 172 -KEQDKEM 179
Db 1617 SKEELKAL 1625
RESULT 20
NCR2_HUMAN
ID NCR2_HUMAN STANDARD; PRT; 2517 AA.
AC Q9Y618; Q9Y500; Q13354; O00613; O15416;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclear receptor co-repressor 2 (N-CoR2) (Silencing mediator of
DE retinoic acid and thyroid hormone receptor) (SMRT) (SMRte) (Thyroid-,
DE retinoic-acid-receptor-associated co-repressor) (T3 receptor-
DE associating factor) (TRAC) (CTG26).
GN NCR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM SMRT).
TISSUP=Pituitary;
MEDLINE=99178941; PubMed=10077563;
RA Ordenlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;
RA "Unique forms of human and mouse nuclear receptor corepressor SMRT.";
RA Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
[2]
SEQUENCE FROM N.A. (ISOFORM SMRT).
TISSUP=Cervical adenocarcinoma;
MEDLINE=9919215; PubMed=10097068;
RA Park E.J., Schroeen D.J., Yang M., Li H., Li L., Chen J.D.;
RA "SMRte, a silencing mediator for retinoid and thyroid hormone
RA receptors-extended isoform that is more related to the nuclear
RT
receptor corepressor.";
Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
[3]
SEQUENCE OF 1023-2517 FROM N.A.
TISSUP=Cervical adenocarcinoma;
MEDLINE=96008552; PubMed=7566127;
RA Chen J.D., Evans R.M.;
RA "A transcriptional co-repressor that interacts with nuclear hormone
RT receptors.";
Nature 377:454-457(1995).
[4]
SEQUENCE FROM N.A. (ISOFORM TRAC-1).
TISSUP=Fetal liver;
MEDLINE=96408715; PubMed=8813722;
RA Sande S., Privalsky M.L.;
RA "Identification of TRACs (T3 receptor-associating cofactors), a family
RT of cofactors that associate with, and modulate the activity of,
RT nuclear hormone receptors.";
Mol. Endocrinol. 10:813-825(1996).
[5]
SEQUENCE OF 428-613 FROM N.A.
TISSUP=Brain cortex;
MEDLINE=97369492; PubMed=9225980;
RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RA "cDNAs with long CAG trinucleotide repeats from human brain.";
Hum. Genet. 100:114-122(1997).
CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTOR (RAR) IN THE
CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIIB.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SMRT/TRAC-2 (SHOWN HERE) AND
CC TRAC-1 ARE PRODUCED BY ALTERNATIVE SPLICING. TRAC-1 CONTAINS ONLY
CC THE C-TERMINAL RECEPTOR-INTERACTING DOMAIN AND ACTS AS AN
CC ANTIREPRESSOR.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS. HIGH LEVELS OF EXPRESSION ARE
CC DETECTED IN LUNG, SPLEEN AND BRAIN.
CC -1- INDUCTION: REGULATED DURING CELL CYCLE PROGRESSION.
CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
CC DOMAINS (ID1 AND ID2).
CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
CC SPECIFICITY.
CC -1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CORNR BOXES.
CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF113003; AAD20946.1; -.
DR EMBL; AF125672; AAD22973.1; -.
DR EMBL; U37116; AAC50236.1; -.
DR EMBL; S83390; AAB50847.1; -.
DR EMBL; U80750; AAB91446.1; -.
DR TRANSFAC; T04689; -.
DR Genew; HGNC:7673; NCR2.

```



"Novel members of the cdc2-related kinase family in Drosophila: cdk4/6, cdk5, PFTAIKE, and PITSLRE kinase.";

Mol. Biol. Cell 7:1759-1769(1996).

[2]

SEQUENCE FROM N.A.

STRAIN-Berkeley:

MEDLINE=20196006; PubMed=10731132;

Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D., Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W., Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of *Drosophila melanogaster*.";

Science 287:2185-2195(2000).

[3]

SEQUENCE OF 571-722 FROM N.A.

TISSUE=Imaginal disks;

MEDLINE=92335284; PubMed=1378625;

Biggs W.H. III, Zipursky S.L.;

"Primary structure, expression, and signal-dependent tyrosine phosphorylation of a *Drosophila* homolog of extracellular signal-regulated kinase.";

Proc. Natl. Acad. Sci. U.S.A. 89:6295-6299(1992).

-!- FUNCTION: Acts as a negative regulator of the normal cell cycle progression. May function in regulating proliferation by the phosphorylation and subsequent plasma membrane targeting of galactosyltransferase (By similarity).

-!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.

-!- SUBCELLULAR LOCATION: Nuclear (Potential).

-!- TISSUE SPECIFICITY: Present throughout the early embryo. In late embryos levels are highest in the CNS.

-!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically. Highest levels in early embryogenesis (0-6 hours), low levels during later embryogenesis, moderate levels in pupae and adults.

-!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. CDC2/CDCX SUBFAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X99513; CAA67863.1; -;

DR EMBL; AE003592; AAF51635.1; -;

DR HSP; P24941; 1B38.

DR FlyBase; FBgn0016696; Pitslre.

DR InterPro; IPR000719; Euk\_pkinase.

DR InterPro; IPR002290; Ser\_thr\_pkinase.

DR Pfam; PF00069; pkinase; 1.

DR ProDom; PD000001; Euk\_pkinase; 1.

DR SMART; SM00220; S\_TKc; 1.

DR PROSITE; PS0107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.

DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.

DR Transferase; Serine/threonine-protein kinase; ATP-binding; Cell cycle; Nuclear protein.

KW DOMAIN 1 51 56 NUCLEAR LOCALIZATION SIGNAL

FT DOMAIN 92 146 (POTENTIAL).

FT DOMAIN 1386 504 HIS-RICH.

FT DOMAIN 558 851 SER-RICH.

FT NP\_BIND 564 572 PROTEIN KINASE.

FT BINDING 587 587 ATP (BY SIMILARITY).

FT ACT\_SITE 685 685 BY SIMILARITY.

FT CONFLICT 69 69 D -> E (IN REF. 1).

FT CONFLICT 223 223 V -> A (IN REF. 1).

FT CONFLICT 283 283 I -> S (IN REF. 1).

FT CONFLICT 492 492 V -> A (IN REF. 1).

FT CONFLICT 584 584 V -> E (IN REF. 3).

FT CONFLICT 603 603 L -> R (IN REF. 3).

FT CONFLICT 612 612 G -> GG (IN REF. 3).

FT CONFLICT 670 670 MISSING (IN REF. 3).

FT CONFLICT 680 680 MISSING (IN REF. 3).

FT CONFLICT 717 717 MISSING (IN REF. 3).

SQ SEQUENCE 952 AA; 108837 MW; 9CBDE8D459D0713D CRC64;

Query Match 9.0%; Score 90.5; DB 1; Length 952;

Best Local Similarity 22.18; Pred. No. 5.8;

Matches 42; Conservative 35; Mismatches 58; Indels 45; Gaps 6;

QY 21 IHFEN:VKSDIDFQVLVINEKSAFDVTVFGQRFSEILLKYDF--IVGDWNGEQLRLRGFY 73

DB 350 IHVRRKSKPDNYEIKLKRRREDDIEVIRDDDDSESDSESDNEEVPPQDSEGSATESGS 409

QY 79 KDASTIRKNSRI---SRLEDIKEYCNFGCAYFVLENPNPRIKFD--ERPH 126

DB 410 EDSYA:SKKSKIKSKSQLED-----DDEDLPDPSPLSVGLYKSPK 451

QY 127 KRRKS:SKSQSSQSQ--TRNRSQSNANAHETSKRKD-----TKRROERHIKE 173

DB 452 QRQRS:SVSSKSSQSSKSSRSRSRSRSQSSLEDEVDQVGDADSPSSSTRSERGWTQ 511

QY 174 EQDKETSAK 183

DB 512 EQPEEKPEEK 521

RESULT 18

APU\_THETU ID APU\_THETU STANDARD; PRT; 1861 AA.

AC P38536;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Anylopullulanase precursor (Alpha-amylase/pullulanase) (pullulanase type II) [includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase); pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase) (Alpha-dextrin endo-1,6-alpha-glucosidase)].

GN AMYB.

OS Thermoanaerobacter thermosulfurogenes (Clostridium thermosulfurogenes).

OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;



```

FT DOMAIN 466 574 ARG/SER-RICH.
FT DOMAIN 664 814 ARG/SER-RICH.
FT DOMAIN 1311 1348 ARG-SER TANDEM REPEAT-RICH.
SQ SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;

Query Match
Best Local Similarity 9.1%; Score 91.5; DB 1; Length 1462;
Matches 32; Conservative 14; Mismatches 41; Indels 27; Gaps 4;

QY 78 YKD---ASTIRKNSRLSRLDIYKVCNFCGAYFVLENPNPRDIKFDDEPHKRRKSRK 134
DB 681 YSDRSESSPRKSRSSRSRSYR-----SRSLASSHSR 720
QY 135 SOSKSSOTRN---NRSSQANAHFTSKRKDKRROERHIKEQDKEMTSKOH 185
DB 721 SPSRSHSNKYSDHSQCRSSSYTIS-SDGGRAKRLRSGKNSVSHKX 773

RESULT 14
SFR6.HUMAN
ID SFR6.HUMAN STANDARD; PRT: 344 AA.
AC Q13247; Q13244; Q13245; Q9UJB8; Q96J06;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Splicing factor, arginine/serine-rich 6 (Pre-mRNA splicing factor
DE SRP55)
GN SFR56 OR SRP55.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Colon;
RX MEDLINE=96016206; PubMed=7556075;
RA Screation G.R., Caceres J.F., Mayeda A., Bell M.V., Plebanski M.,
RA Jackson D.G., Bell J.I., Krainer A.R.;
RT "Identification and characterization of three members of the human SR
RT family of pre-mRNA splicing factors.";
RL EMBO J. 14:4336-4349(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharvalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A.C., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM SRP55-1).

```

```

RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE ON 21-27 AND 47-55.
RX MEDLINE=92:49775; PubMed=1577277;
RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
RT "SR proteins: a conserved family of pre-mRNA splicing factors.";
RL Genes Dev. 16:837-847(1992).
CC -!- FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE
CC THE SELECTION OF ALTERNATIVE SPLICING SITES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; SRP55-1 (SHOWN HERE), SRP55-
CC 2 AND SRP55-3; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -!- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: U30833; AAA93073.1; .
DR EMBL: U30828; AAA93071.1; .
DR EMBL: U30829; AAA93072.1; .
DR EMBL: AL031681; CAB43960.1; .
DR EMBL: BC005832; AAB06832.1; .
DR Genew: HGN2:10788; SFR56.
DR MIM: 601944; .
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS01102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
KW Repeat; Phosphorylation.
FT DOMAIN 1 72
FT DOMAIN 87 90 RNA-BINDING (RRM) 1.
FT DOMAIN 110 183 GLY-RICH (HINGE REGION).
FT DOMAIN 184 343 RNA-BINDING (RRM) 2.
FT VARSPPLIC 86 135 ARG/SER-RICH (RS DOMAIN)
FT VARSPPLIC 136 344 SGGGYSRRSRSGDKYGPVPTVTEARLIVENLSSRCSWDL
FT VARSPPLIC 313 344 KDFMRQAGE -> MTNGAEAVSTEAKMTAFDPDWLFFHFLC
FT VARSPPLIC 64 64 DPCPMTLTLTLPEAMTTAAFC (IN ISOFORM SRP55-
FT VARSPPLIC 344 344 2).
FT CONFLICT 64 64 MISSING (IN ISOFORM SRP55-2).
FT CONFLICT 344 344 RSVPPKRRATSRSRSRSRSRSSSRD -> LKLGA
FT CONFLICT 64 64 RFMSQQTESIYSLASSC (IN ISOFORM SRP55-3).
SQ SEQUENCE 344 AA; 39586 MW; 72305506CE948B94 CRC64;

Query Match
Best Local Similarity 9.0%; Score 91; DB 1; Length 344;
Matches 52; Conservative 34; Mismatches 83; Indels 56; Gaps 9;

QY 17 GPQFTHFENIVKSDDIEFQVLVINEKSAFDVTV---FGQRFSEILLKYDFIVGDWNEQLR 73
DB 36 GYGFVEFEDSRDADDAVVELNGKELCGERVIVEHARGPRDRDGYSGSRSGGGYSSRR 95
QY 74 LRGFVKDASTIKNSRI-----SR-----LEDYI-----KEYCNFGCAYF-- 108
DB 96 TSGRIKYGPPVTEYRLIVENLSSRCSWDLKDFMRQAGEVTYADAHKRTNEGVTEFRS 155
QY 109 -----VLE-----NPNPRDIKFDDEP-----HKRRKSRKS-QSSKSO 141
DB 156 YSDMIRALDKLDGTEINGRNIRLIEDKPTSHRRSYGSRSRSRSRSRSSRSR 215
QY 142 TRN-IRSSQSNANAHFTSKRKDKTKRQERHIKEQDKEMTSKOH 185

```









```

[3] X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 1-177.
RC STRAIN=ATCC 75593;
RX MEDLINE=98318040; PubMed=9655334;
RA Mikol V., Ma D., Carlow C.K.S.;
RT "Crystal structure of the cyclophilin-like domain from the parasitic
RT nematode Brugia malayi.";
RL Protein Sci. 7:1310-1316(1998).
[4]
RN RP X-RAY CRYSTALLOGRAPHY (2.47 ANGSTROMS) OF 1-177.
RX MEDLINE=20108543; PubMed=10642184;
RA Ellis P.J., Carlow C.K.S., Ma D., Kuhn P.;
RT "Crystal structure of the complex of Brugia malayi cyclophilin and
RT cyclosporin A.";
RL Biochemistry 39:592-598(2000).
CC -!- FUNCTION: PPIases accelerate the folding of proteins. It catalyzes
CC the cis-trans isomerization of proline imidic peptide bonds in
CC oligopeptides.
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- ENZYME REGULATION: RELATIVELY INSENSITIVE TO INHIBITION BY CSA.
CC -!- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L37292; AAC37249.1; -.
DR PDB; 1A58; 27-MAY-98.
DR PDB; 1A33; 29-JUL-98.
DR PDB; 1C5F; 02-DEC-99.
DR InterPro; IPR002130; CSA_PPIase.
DR Pfam; PF00160; pro-isomerase; 1.
DR PRINTS; PR00153; CSAPPISMRASE.
DR PROSITE; PS00170; CSA_PPIASE_1; 1.
DR PROSITE; PS00072; CSA_PPIASE_2; 1.
DR Isomerase; Rotamase; 3D-structure.
KW DOMAIN 10 175 PPIASE, CYCLOPHILIN-TYPE.
FT DOMAIN 700 709 POLY-ARG.
FT DOMAIN 713 716 POLY-ARG.
FT DOMAIN 800 815 POLY-SER.
FT DOMAIN 828 837 POLY-SER.
SQ SEQUENCE 843 AA; 97817 MW; 3C34EC90A32EDBDC CRC64;
Query Match 9.6%; Score 96.5; DB 1; Length 843;
Best Local Similarity 21.9%; Pred. No. 1.7;
Matches 41; Conservative 37; Mismatches 56; Indels 53; Gaps 8;
QY 28 KSDIDIEFOLVINEKSAFDVTFVGORFSE-----ILLKYDFIVGDWGNQLRGFYKDASTIR 85
DB 603 KLDDLKRKETSGQKQAD-----SEQIVAEKTNVVDNSDNSSKMSVNGKLEKVESTN 654
QY 86 KNSRISRKLEDY-----IKYCNFGCAIFYLE-----NPNPRDIKFDDBRPKHRR 129
DB 655 KENEVSQKDLKAESTKSEIKQOVN-----EVSRRKQKGKPKHEKRNRSRRRR 706
QY 130 KS-----RSKSSSKSOTFNNRNSOSNANHFT-----SKKRKDKTR-----RQERH 170
DB 707 RSRNSGRRRRSSRRSRDRRHKRSRSGYVRRFEGWSRRRPTRELYDERMRRE 766
QY 171 IKKEQDK 177
DB 767 RRSSEDR 773
;
RESULT 10
APU_THESA STANDARD; PRT; 1279 AA.
ID APU_THESA
AC P36905;

```

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>), or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M74002; AAA35554.1; -;  
 Genew; HGNC:10782; SFRS11.  
 MIM; 602010; -;  
 InterPro; IPR000504; RNA\_rec\_mot.  
 Pfam; PF00076; rrm; 1.  
 SMART; SM00360; RRM; 1.  
 PROSITE; PS0102; RRM; 1.  
 PROSITE; PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 Nuclear protein; RNA-binding; mRNA splicing; Repeat.  
 FT DOMAIN 17 31  
 POLY-GLY.  
 FT DOMAIN 33 113  
 RNA-BINDING (RRM).  
 FT DOMAIN 252 259  
 POLY-ARG.  
 FT DOMAIN 291 294  
 POLY-ARG.  
 FT DOMAIN 341 349  
 POLY-ARG.  
 FT DOMAIN 397 404  
 POLY-LYS.  
 FT DOMAIN 247 353  
 10 X 8 AA APPROXIMATE R-R-S-R-S-R-S-R  
 REPEATS.  
 FT REPEAT 247 255  
 1.  
 FT REPEAT 258 265  
 2.  
 FT REPEAT 267 274  
 3.  
 FT REPEAT 275 282  
 4.  
 FT REPEAT 285 292  
 5.  
 FT REPEAT 293 300  
 6.  
 FT REPEAT 302 309  
 7.  
 FT REPEAT 321 328  
 8.  
 FT REPEAT 334 341  
 9.  
 FT REPEAT 346 353  
 10.  
 SQ SEQUENCE 484 AA; 53542 MW; 1211E96DDCOA3182 CRC64;

Query Match 9.8%; Score 99; DB 1; Length 484;

Best Local Similarity 34.3%; Pred. No. 0.57;

Matches 24; Conservative 17; Mismatches 21; Indels 8; Gaps 2;

Oy 122 DERPHKRKRSKSSQS-----SKSQTNRNRSQSNANAHFTSKRKDKTK--RRQERHIKE 173  
 Db 240 DKKEKRRHRSKSRSSRRRTPTSSRRHRSKSRSSRRSHSKSRSSRRSHSHSRE 299  
 Oy 174 EODKEMTSK 183  
 Db 300 RGRRSKSTSK 309

RESULT 7

SFR5\_MOUSE  
 ID SFR5\_MOUSE STANDARD; PRT; 270 AA.  
 AC O35326;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Splicing factor, arginine/serine-rich 5 (Pre-mRNA splicing factor SRP40) (Delayed-early protein HRS).  
 GN SFR5 OR HRS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=98094279; PubMed=9434190;  
 RA Du K., Taub R.;  
 RT "Alternative splicing and structure of the human and mouse SFR5/HRS/SRP40 genes.";  
 RL Gene 204:243-249(1997).  
 CC -!- FUNCTION: MAY BE REQUIRED FOR PROGRESSION THROUGH G1 AND ENTRY INTO S PHASE OF CELL GROWTH. MAY PLAY A REGULATORY ROLE IN PRE-MRNA SPLICING. AUTOREGULATES ITS OWN EXPRESSION. PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE THE SELECTION OF ALTERNATIVE SPICE SITES (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

-!- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS DOMAIN (BY SIMILARITY).  
 -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).  
 -!- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: AF020308; AAC39946.1; -;  
 MGD; MGI:98287; Sfrs5.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 2.  
 DR SMART; SM00360; RRM; 2.  
 DR PROSITE; PS0102; RRM; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 Transcripton regulation; Nuclear protein; RNA-binding; mRNA splicing;  
 KW DOMAIN 4 74  
 Repeat; Phosphorylation.  
 FT DOMAIN 78 83  
 RNA-BINDING (RRM) 1.  
 FT DOMAIN 108 190  
 RNA-BINDING (RRM) 2.  
 FT DOMAIN 191 265  
 ARG/SER-RICH (RS DOMAIN).  
 SQ SEQUENCE 270 AA; 30944 MW; 277397C3DC657895 CRC64;  
 Query Match 9.8%; Score 98.5; DB 1; Length 270;  
 Best Local Similarity 20.9%; Pred. No. 0.33;  
 Matches 44; Conservative 46; Mismatches 68; Indels 53; Gaps 8;  
 Oy 20 FIHEHIVKSDDIEFQVLVINEKSAFDVTV-----FCQ-RPESEILLKYDFIVGQG 63  
 Db 41 FVEFEDPRDADAVYELDGKELCSERVITIEHARARSGRGRGYSDDL--G--AVADLG 95  
 Oy 69 NEOLR-----LRGFYKDASTIR-KNSRISRLEDYIKEYCNFG 104  
 Db 96 NDRKK-STCKNRNLIVENLSRVSWQDLKDFMQAGEVTFADHRPKLNGVVEFAFYG 155  
 Oy 105 CAYEVLENPNPRDIFDDEP-----HKRRKSRKSSQS-----SKSQTNRNRSQSNAN 152  
 Db 156 DLKNA[EKLWKR-KLTGEKSKLIEGSKRHSRSRSRSTRSSLSRSTRSRSSRSKYSR 214  
 Oy 153 AHFTS[KRKDTKRQERHIKEEQDKEMTSK 183  
 Db 215 SRSRSRSRSKSRSGSRSPVPEKSKQKRGSSR 245  
 RESULT 8  
 RA50\_SULAC  
 ID RA50\_SULAC STANDARD; PRT; 886 AA.  
 AC O33600;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA double-strand break repair rad50 ATPase.  
 GN RAD50.  
 OS Sulfolobus acidocaldarius.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OX NCBI\_TaxID=2285;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;  
 RX MEDLINE=97362314; PubMed=9211741;  
 RA Elie C., Baucher M.F., Fondrat C., Forterre P.;  
 RT "A protein related to eucaryal and bacterial DNA-motor proteins in the hyperthermophilic archaeon Sulfolobus acidocaldarius.";  
 RL J. Mol. Evol. 45:107-114(1997).  
 CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The rad50/mre11 complex possesses single-strand endonuclease activity and ATP-dependent double-strand-specific exonuclease activity.

```

QY 156 TSKRRKDTKRQERHKEQDKEMTSK 183
D 156 TSKRRKDTKRQERHKEQDKEMTSK 183
D 217 RSRSRKSRSGRSRSPVPEKSKRGSSSR 244
D 217 RSRSRKSRSGRSRSPVPEKSKRGSSSR 244

RESULT 5
SFRR_HUMAN STANDARD; PRT; 272 AA.
AC Q13243; Q16662; O14797;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor, arginine/serine-rich 5 (pre-mRNA splicing factor
DE SRP40) (Delayed-early protein HRS).
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=96016206; PubMed=7556075;
RA Sreaton G.R., Caceres J.F., Mayeda A., Bell M.V., Plebanski M.,
RA Jackson D.G., Bell J.R., Krainer A.R.;
RT "Identification and characterization of three members of the human SR
RT family of pre-mRNA splicing factors.";
RL EMBO J. 14:4336-4349(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Skin;
RA Strausberg R.;
RX Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-192 FROM N.A.
RX MEDLINE=98094279; PubMed=9434190;
RA Du K., Taub R.;
RT "Alternative splicing and structure of the human and mouse
RT SFRS5/HRS/SRP40 genes.";
RL Gene 204:243-249(1997).
RN [4]
RP SEQUENCE OF 130-158.
RX MEDLINE=92249775; PubMed=1577277;
RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
RT "SR proteins: a conserved family of pre-mRNA splicing factors.";
RL Genes Dev. 6:837-847(1992).
CC -1- FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE
CC THE SELECTION OF ALTERNATIVE SPLICE SITES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; SRP40-1 (SHOWN HERE), SRP40-
CC 2, SRP40-3 AND SRP40-4; ARE PRODUCED BY ALTERNATIVE SPLICING AND
CC DIFFER IN THEIR C-TERMINAL DOMAINS.
CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U30826; AAA93070.1;
CC EMBL; U30884; AAA93074.1;
CC EMBL; U30827; AAB60405.1;
CC EMBL; BC018823; AAU18823.1;
CC EMBL; AF020307; AAC39543.1;
CC Genew; HGNC:10787; SFRS5.

MIM; 600914;
InterPro: IPR000504; RNA_rec_mot.
Pfam: PF00076; rrm; 2.
SMART: SM00360; RRM; 2.
PROSITE: PS0102; RRM; 2.
PROSITE: PS00030; RRM_RNP_1; 1.
KW Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
KW Repeat; Phosphorylation.
FT DOMAIN 4 74 RNA-BINDING (RRM) 1.
FT DOMAIN 78 83 GLY-RICH (HINGE REGION).
FT DOMAIN 108 181 RNA-BINDING (RRM) 2.
FT DOMAIN 182 267 ARG/SER-RICH (RS DOMAIN).
FT VARSPPLIC 43 45 MISSING (IN ISOFORM SRP40-4).
FT VARSPPLIC 100 107 NAPPVTE -> YVKGGLW (IN ISOFORM SRP40-
FT 2).
FT VARSPPLIC 108 272 MISSING (IN ISOFORM SRP40-2).
FT VARSPPLIC 113 140 ENLSRVSWQDLKDFMRQAGVTFADAH -> GEFILKSQL
FT AGSQRFHETSWGNSVCGCT (IN ISOFORM SRP40-4).
SQ SEQUENCE 272 AA; 31263 MW; F13AD79845ECBB16 CRC64;

Query Match 9.9%; Score 99.5; DB 1; Length 272;
Best Local Similarity 20.4%; Pred. No. 0.28;
Matches 43; Conservative 44; Mismatches 73; Indels 51; Gaps 8;

QY 20 FIHPENIVKSDIEFQVLVINEKSAFDVTV-----FGQFPSEILLYD--- 61
D 20 FIHPENIVKSDIEFQVLVINEKSAFDVTV-----FGQFPSEILLYD--- 61
D 41 FVEFEDPRDAVDVYELDGKELCSERVIEHARARSRGGRGGRYSDRFSRRPRNDRN 100
QY 62 -----FIVGD-----WGNEQLRLRGFYKDASTIR-KNSRISRLDYIKEYCNFGC 105
D 62 -----FIVGD-----WGNEQLRLRGFYKDASTIR-KNSRISRLDYIKEYCNFGC 105
D 101 APPVRTENRLIVENLSRVSWQD-----LKDFMRQAGVTFADAHRLKLNQGVVEFASIGD 156
QY 106 AYEVL-----NPNPRDIK-FDDERPHKRRKSRKSKS-----SKSQTRNNRNSQSNAN 152
D 106 AYEVL-----NPNPRDIK-FDDERPHKRRKSRKSKS-----SKSQTRNNRNSQSNAN 152
D 157 LKNAIEKLSCGKEINGRKIKLEGRHSRKSRSRSTRSSRSRKSRSRKSRSRKSRSR 216
QY 153 AHTSKRKDTKRQERHKEQDKEMTSK 183
D 153 AHTSKRKDTKRQERHKEQDKEMTSK 183
D 217 RSRSRKSRSGRSRSPVPEKSKRGSSSR 247

RESULT 6
SFRR_HUMAN STANDARD; PRT; 484 AA.
AC Q05519;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor arginine/serine-rich 11 (Arginine-rich 54 kDa nuclear
DE protein) (p54).
GN SFRS11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hepatoma;
RX MEDLINE=91376109; PubMed=1896467;
RA Chaudhary N., McMahon C., Blobel G.;
RT "Primary structure of a human arginine-rich nuclear protein that
RT colocalizes with spliceosome components.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8189-8193(1991).
CC -1- FUNCTION: MAY FUNCTION IN PRE-MRNA SPLICING.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZES WITH SPliceosome
CC COMPONENTS.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```
CC -----
DR EMBL; L14076; AAA36649.1; -.
DR EMBL; BC002781; AA020781.1; -.
DR EMBL; AC004236; AAC04476.1; -.
DR Genew; HGNC:10786; SFRS4.
DR MIM; 601940; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS00102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW mRNA processing; mRNA splicing; Nuclear protein; RNA-binding;
KW Repeat; Phosphorylation.
FT DOMAIN 2 72 RNA-BINDING (RRM) 1.
FT DOMAIN 72 97 GLY-RICH (HINGE REGION).
FT DOMAIN 104 177 RNA-BINDING (RRM) 2.
FT DOMAIN 179 494 ARG/SER-RICH (RS DOMAIN).
FT CONFLICT 35 35 N -> D (IN REF. 4).
FT CONFLICT 253 253 E -> D (IN REF. 1).
FT CONFLICT 318 322 SGRS -> EQQE (IN REF. 1).
FT CONFLICT 338 338 G -> A (IN REF. 1).
FT CONFLICT 356 356 G -> S (IN REF. 1).
FT CONFLICT 436 438 TNQ -> RNE (IN REF. 1).
SQ SEQUENCE 494 AA; 56678 MW; 5BBAB917C218C20A CRC64;

Query Match 10.1%; Score 102; DB 1; Length 494;
Best Local Similarity 22.4%; Pred. No. 0.34;
Matches 52; Conservative 35; Mismatches 75; Indels 70; Gaps 10;

Qy 17 GPQIFHNIVKSDIEFOLVINEKSAFDVTV-----FGQRFSSILLKYD--- 54
Db 36 GYGFVEFDLRDADDAVYELNGKDLGGERVIVEHARGPRDGSYSGRSGYRRSGRDK 95
Qy 55 ---EILLKYDIVGD-----WGN--EQLRLG--FYKDASTIRKNSIRLEDYKEYC 101
Db 96 YGPPTRTYRLIVENLSRCSWQDLKDYMQAGEVTVYADAHKGRN-----EGVIEFV 148
Qy 102 NFGCAYFYLE-----NPNPRDIKFDDRP-----HKR-----RKSRKS 135
Db 149 SYSOMKRALEKLDGTEVNGRKIRLVEDKPGRRRSYSRSHSRSHSRSHSRSR 208
Qy 136 QSSKSTQNNRSQSNANAHFTSK--KRKDTKRQERHTEKQDKEMTSAKOH 185
Db 209 GSSKSSHSKRSRSGSRSRKSRKSRKSRKSRKSRKSRKSRKSRKSRSH 260

RESULT 4
SFR5_RAT
ID SFR5_RAT STANDARD: PRT; 269 AA.
AC Q09167; Q35335;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Splicing factor, arginine/serine-rich 5 (Pre-mRNA splicing factor
DE SRP40) (Insulin-induced growth response protein CL-4) (Delayed-early
DE protein HRS).
GN SFRS5 OR CL-4 OR HRS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=93315501; PubMed=7686911;
RA Diamond R.H., Du K., Lee V.M., Mohn K.L., Haber B.A., Tewari D.S.,
RA Taub R.;
RT "Novel delayed-early and highly insulin-induced growth response
RT genes. Identification of HRS, a potential regulator of alternative
RT pre-mRNA splicing."
RL J. Biol. Chem. 268:15195-15192(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
```

```
RC TISSUE=Thymus;
RX MEDLINE=98094279; PubMed=9434190;
RA Du K., Taub R.;
RT "Alternative splicing and structure of the human and mouse
RT SFRS5/HRS/SRP40 genes."
RL Gene 204:243-249(1997).
RN [3]
RP SEQUENCE OF 185-269 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=94164020; PubMed=8161377;
RA Hamil K.G., Hall S.H.;
RT "Cloning of rat Sertoli cell follicle-stimulating hormone primary
RT response complementary deoxyribonucleic acid: regulation of TSC-22
RT gene expression."
RL Endocrinology 134:1205-1212(1994).
CC -1- FUNCTION: MAY BE REQUIRED FOR PROGRESSION THROUGH G1 AND ENTRY
CC INTO S PHASE OF CELL GROWTH. MAY PLAY A REGULATORY ROLE IN PRE-
CC MRNA SPLICING. AUTOREGULATES ITS OWN EXPRESSION. PLAYS A ROLE IN
CC CONSTITUTIVE SPLICING AND CAN MODULATE THE SELECTION OF
CC ALTERNATIVE SPICE SITES (BY SIMILARITY).
CC -1- FUNCTION: COULD PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC DIFFERENTIATION IN THE SPLEEN AND THYMUS.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2/HRR-LF; are
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SPLEEN AND THYMUS.
CC -1- INDUCTION: BY INSULIN AND HEPATECTOMY.
CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L13615; AAA62266.1; -.
DR EMBL; AF020683; AAB1864.1; -.
DR EMBL; L32567; AAA42316.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS00102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR Transcription regulation; Nuclear protein; RNA-binding; mRNA splicing;
KW Repeat; Phosphorylation; Alternative splicing.
FT DOMAIN 4 74 RNA-BINDING (RRM) 1.
FT DOMAIN 75 107 ARG/GLY-RICH (HINGE REGION).
FT DOMAIN 108 181 RNA-BINDING (RRM) 2.
FT DOMAIN 184 264 ARG/SER-RICH (RS DOMAIN).
FT VARSPIC 123 124 DL -> VC (IN ISOFORM 2).
FT VARSPIC 125 269 MISSING (IN ISOFORM 2).
FT CONFLICT 69 69 I -> M (IN REF. 2).
SQ SEQUENCE 1269 AA; 30891 MW; 0FC90628B64DE845 CRC64;

Query Match 9.9%; Score 100; DB 1; Length 269;
Best Local Similarity 20.7%; Pred. No. 0.25;
Matches 43; Conservative 43; Mismatches 74; Indels 48; Gaps 7;

Qy 20 FIHEENIVKSDIEFOLVINEKSAFDVTV-----FGQRFSSILLKYD--- 61
Db 41 FVEFEDPDADDAVYELDGKELCSERVVIEHARSRGGRGGRGYSRDRFSRRPRNDRN 100
Qy 62 -----FTVGD-----WGNQLRLRGYKDASTR-KNSRISRLEDYKEYCNFGC 105
Db 101 APPVTEENRLIVENLSRVSQD---LKDFMRQAGEVTFADAHKPKLNEGVEEASGVD 156
Qy 106 AYFVLE-----NPNPRDIKFDDERPHKRSKRSQS-----SKSQTNNRSQSNANAHF 155
Db 157 LKNAIEKLSGKINGRKIKLEGSRKRSRSTRSRSRSRSRSRSRSRSRSR 216
```





```
Db 1128 SNKKKEDKNEKKKSQHVKLKVK 1149
RESULT 39
Q8REF7
AC Q8REF7 PRELIMINARY; PRT; 1056 AA.
AC Q8REF7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created).
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ATP-dependent nuclease subunit A.
GN FN1149.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010620; AAL95345.1; -.
KW Complete proteome.
SQ SEQUENCE 1056 AA; 125480 MW; D07C14C2AC6F6465 CRC64;

Query Match 9.0%; Score 90.5; DB 16; Length 1056;
Best Local Similarity 22.5%; Pred. No. 27;
Matches 48; Conservative 41; Mismatches 67; Indels 57; Gaps 11;

QY 8 EMLNKKYGPQF-----IHFENIVKSDIEFOLVINEKSAFDVTF-----GQR--- 52
Db 736 ELVTIHKSGLEFKFTTVINKSKKSTDDIDFLFEMNDK--YDKTVSFLCKKGKYPILK 793
QY 53 --FSEILLKYDFIVGDMGNEQLRGYKDASTIRKNSRISLEDYIKCYNFGCAYFVL 110
Db 794 TCPERIENYDKKI-----KEEENINFYV-ALTRPKNNLIVIEDRLF----- 836
QY 111 ENP-----NPRD-----TKFDDERPHKRKRKRSK---SQSKSQTRNNRSQSNANAHF 155
Db 837 ENPLNESNIDDFNCELGKISLDE-----KSKTEDIIKKNLENDLYNSQSVFSSSIYE 890
QY 156 TSKKRKDTKRQRERHKEEODKEMTSAKQHLF 188
Db 891 NEEIKNIEVNESKFLLETEEKRMIGILVHYFF 923

RESULT 40
Q9UHV3
AC Q9UHV3 PRELIMINARY; PRT; 1103 AA.
AC Q9UHV3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pancreas-enriched phospholipase C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawasaki H., Chen E.J., Springett G.M., Graybiel A.M., Housman D.E.;
RT "A novel phospholipase C enriched in pancreas.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF117948; AAF22005.1; -.
DR HSSP; P10688; 1DJX.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001192; PI_PLC.
```

```
DR InterPro; IPR000909; PI_PLC_xdom.
DR InterPro; IPR001711; PI_PLC_Y.
DR InterPro; IPR000159; RA_domain.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00388; PI-PLC-X; 1.
DR Pfam; PF00387; PI-PLC-Y; 1.
DR Pfam; PF00788; RA; 1.
DR PRINTS; PR00390; PHPLIPASEC.
DR ProDom; PD001202; PI_PLC_Y; 1.
DR SMART; SM00239; C2; 1.
DR SMART; SM00148; PLCXC; 1.
DR SMART; SM00149; PLCYC; 1.
DR SMART; SM00314; RA; 1.
DR PROSITE; PS50004; C2_DOMAIN_2; 1.
DR PROSITE; PS50007; PIPLC_X_DOMAIN; 1.
DR PROSITE; PS50008; PIPLC_Y_DOMAIN; 1.
SQ SEQUENCE 1103 AA; 124905 MW; 29BC730DB4377995 CRC64;

Query Match 9.0%; Score 90.5; DB 4; Length 1103;
Best Local Similarity 24.0%; Pred. No. 28;
Matches 46; Conservative 21; Mismatches 62; Indels 63; Gaps 7;

QY 24 ENIVKSDDIEFOLVINEKSAFDVTFVFGQRFSEILLKYDFIVGDMGNEQLRLRGFYKDAST 83
Db 920 KNIVQ-DDKEVILSSEESFF-----VQVHDSVPEQPR-----T 952
QY 84 IRKNSRISLEDYIKE-YCNFGCAYFVLENPNPRDIKEDDE----RPHKRRKRSKSKSQ-- 136
Db 953 VIRAPRVSTAQDVIQQTLCRAKYSYLSNPNPSDYVLLLEEVVKDNTTKKTTTPKSSQRV 1012
QY 137 -----SSKSQTRNNRSQSNANAHFTSKKRKDTKRQRERHKEE 174
Db 1013 LLDQECVFQAQSKWKAGKGFILKKEQVQASREDKKGISFASSELKLT-----STK 1065
QY 175 QDKEMTSAKOHL 186
Db 1066 QPRGLTSPSOLL 1077

Search completed: February 26, 2003, 09:54:18
Job time : 36 secs
```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RA Zhao M., Song H., Li N., Peng Y., Han Z., Chen Z.;
RT "A novel gene expressed in human bone marrow.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF208853; AAF64267.1; -
SQ SEQUENCE 173 AA; 20900 MW; 20E3D2AE177DBCC3 CRC64;

Query Match
Best Local Similarity 9.1%; Score 91.5; DB 4; Length 173;
Matches 25; Conservative 17; Mismatches 22; Indels 25; Gaps 3;

Qy 115 PRDIXFDDRPH-----KRRKRSKSSQ-----SKSOTRNNRSOSNANA 153
Db 57 PRSHSYDRRRRRSSSSSYGRRKRSRSRGCRKSYVQRKSRKSRTRRSRPLRS 116

Qy 154 HFTSKK-----RKDKRQRERHIKEEQDK 178
Db 117 HRSKSSRSRRTRRSRDRRRKGRDKE 145

RESULT 37
P71070
ID P71070 PRELIMINARY; PRT; 451 AA.
AC P71070;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-WAR-2002 (TReMBLrel. 20, Last annotation update)
DE YUK[A,B,C,D,E,F], YUK[I,J,K,L,M] and ALD genes.
GN YUKC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Gutseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blandhard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaite Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter K., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrier P., Shin B.S., Soldo B.,
RA Sorokini A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
```

```

subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z820.5; CAB04771.1; -
DR EMBL; Z991.0; CAB15177.1; -
KW Complete proteome.
SQ SEQUENCE 451 AA; 52161 MW; 9F3FFB757377BA2D CRC64;

Query Match
Best Local Similarity 9.1%; Score 91.5; DB 16; Length 451;
Matches 44; Conservative 31; Mismatches 71; Indels 55; Gaps 9;

Qy 4 EVTPEILNLYKPGQFIHFENI--VKSDIDIEFQVIVINEKSAFDTVFGQRFSEILLKYD 61
Db 275 ESLPE:VOVQL--ATSYVEVENLGSAKTKNIENNLVTLSQSDP-----QHFLYWI---- 321

Qy 62 FIVGDNGNEQLRLRGFYKDASTIRKNSRLSLRLEDIKEYCNFGCAFYVL-----EN 112
Db 322 ----D---G-----RGEYKEATSI---GRKLEYN DYI-----YFALAKYKQQLLSDD 359

Qy 113 PNPRD{-----KFDDERPHKRRKRSKSSQSSKSSOTRNNRSOSNANAHTSKRKRD 163
Db 360 TNDED{QKELDSVNSELEKAKQERQENKQSNSETSLVDTSEQTQDDEKQAEKAAEEK 419

Qy 164 KRRQERHIKEEQDKEMTSKQ 184
Db 420 AAAEE{AKKEQKEKEDEKKE 440

RESULT 38
Q9LH98
ID Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2003 (TReMBLrel. 15, Created)
DT 01-OCT-2003 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL; AP002057; BAB03174.1; -
SQ SEQUENCE 2081 AA; D3603EIF85EFFP29 CRC64;

Query Match
Best Local Similarity 9.1%; Score 91.5; DB 10; Length 2081;
Matches 20; Conservative 24; Mismatches 33; Indels 5; Gaps 2;

Qy 111 ENPNFERD{KFDDERPHKRRKRSKSSQSSKSSOTRNNRSOSNANAHTSKRKRDTKRQERH 170
Db 1072 ESENHKSKKEDKKEHEDNKSMKKEDKKEKKHESKSRKK-----EEDKKDKMEKLEQDN 1127

Qy 171 I-KEEQDKEMTSAKOHLFLVRK 191
Db 171 I-KEEQDKEMTSAKOHLFLVRK 191
```



Query Match 9.2%; Score 93; DB 5; Length 925;  
Best Local Similarity 26.5%; Pred. NO. 14;  
Matches 30; Conservative 18; Mismatches 35; Indels 30; Gaps 4;



```
Q9DBP1
ID Q9DBP1 PRELIMINARY; PRT; 339 AA.
AC Q9DBP1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1210001ELIRIK protein.
GN 1210001ELIRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK004831; BAB23599.1; -.
DR MGD: MGI:1915246; 1210001ELIRIK.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS0102; RRM; 2.
SQ SEQUENCE 339 AA; 38997 MW; BC3C9BDA38144A0D CRC64;

Query Match 9.3%; Score 94; DB 11; Length 339;
Best Local Similarity 23.3%; Pred. No. 4;
Matches 50; Conservative 29; Mismatches 88; Indels 48; Gaps 7;

Qy 17 GPQFIHFENIVKSDIEFQVINEKSAFDVTV---FGQFSEILLKYDFIVGDNGNEQLR 73
Db 36 GYGFEFEDSRDADAVYELNSKELCGERVIVEHARGPRDRDGYSGSRGGGYSRR 95

Qy 74 LRGFYKDASTIRKNSRI-----SR-----LEDYI-----KEYCNFGCAV-- 108
Db 96 TSGRDYKGPVPTVEYRLIVENLSRCSQWDLKDFMRQAGEVYADAHKERTNEGVIER 155

Qy 109 -----VLE-----NPNPRDIKFDPERH-----KRRKRSKSSQSSKTSQTNNRSQ 148
Db 156 YSDMKRALDKLDGTEINGRNIRLIEDKPTSHRRSYSGSRSRSRKRSRSRSRSR 215

Qy 149 SNANAHTFSKRRKTKRQERHIKEQDKEMTSK 183
Db 216 SRSTKSRSRSRKSRKSRKSRKSRKSRKSK 250

RESULT 29
Q9LQ35
ID Q9LQ35 PRELIMINARY; PRT; 462 AA.
AC Q9LQ35;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE F14M2.4 protein.

Qy 17 GPQFIHFENIVKSDIEFQVINEKSAFDVTV---FGQFSEILLKYDFIVGDNGNEQLR 73
Db 36 GYGFEFEDSRDADAVYELNSKELCGERVIVEHARGPRDRDGYSGSRGGGYSRR 95

Qy 74 LRGFYKDASTIRKNSRI-----SR-----LEDYI-----KEYCNFGCAV-- 108
Db 96 TSGRDYKGPVPTVEYRLIVENLSRCSQWDLKDFMRQAGEVYADAHKERTNEGVIER 155

Qy 109 -----VLE-----NPNPRDIKFDPERH-----KRRKRSKSKS-QSSKSK 141
Db 156 YSDMKRALDKLDGTEINGRNIRLIEDKPTSHRRSYSGSRSRSRKRSRSRSRSR 215

Qy 142 TRN-NRSQSNANAHTFSKRRKTKRQERHIKEQDKEMTSK 185
```

```
GN F14M2.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsais.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremetska I., Kim C., Lenz C., Li J., Liu S.,
RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC010164; AAF97280.1; -.
SQ SEQUENCE 462 AA; 53790 MW; 708238ED3C11C3C CRC64;

Query Match 9.3%; Score 94; DB 10; Length 462;
Best Local Similarity 26.6%; Pred. No. 5.6;
Matches 42; Conservative 25; Mismatches 63; Indels 28; Gaps 7;

Qy 38 INEKSAFDVTVFGQFSEILLKYDFIVGDNGNEQLRLRGFYKDASTIRKNSRISRLDYI 97
Db 98 ISOREEFTFTLQIFDDKILDY-FIVVFTGGDELE-----ADNQT-----LDYL 142

Qy 98 KEYCNFGCAV---VLENPNPRDIKFDPERHKKRSKSKSQSSKTSQT---RNNRSQSN 150
Db 143 RE----GCPEFLTRVLKCGGRKVLNNKTKDKGKRNQLNQLLAHVTDIRQNGGIPYT 198

Qy 151 ANAHFTSKRRKDKTKRQERHIKEQDKEMTSA--KQHL 186
Db 199 ENMHRIKEENDKLREQESNIDARKKLAEDSAMMKQKL 236

RESULT 30
Q921K3
ID Q921K3 PRELIMINARY; PRT; 339 AA.
AC Q921K3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RKEN CDNA 1210001E11 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012039; AAH12039.1; -.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 2.
DR PROSITE: PS0102; RRM; 2.
SQ SEQUENCE 339 AA; 39029 MW; AE8B6650D0448A0D CRC64;

Query Match 9.2%; Score 93; DB 11; Length 339;
Best Local Similarity 23.1%; Pred. No. 4.9;
Matches 52; Conservative 34; Mismatches 83; Indels 56; Gaps 9;

Qy 17 GPQFIHFENIVKSDIEFQVINEKSAFDVTV---FGQFSEILLKYDFIVGDNGNEQLR 73
Db 36 GYGFEFEDSRDADAVYELNSKELCGERVIVEHARGPRDRDGYSGSRGGGYSRR 95

Qy 74 LRGFYKDASTIRKNSRI-----SR-----LEDYI-----KEYCNFGCAV-- 108
Db 96 TSGRDYKGPVPTVEYRLIVENLSRCSQWDLKDFMRQAGEVYADAHKERTNEGVIER 155

Qy 109 -----VLE-----NPNPRDIKFDPERH-----KRRKRSKSKS-QSSKSK 141
Db 156 YSDMKRALDKLDGTEINGRNIRLIEDKPTSHRRSYSGSRSRSRKRSRSRSRSR 215

Qy 142 TRN-NRSQSNANAHTFSKRRKTKRQERHIKEQDKEMTSK 185
```

```

Db 179 ELFSI:EDGELSDETKKVKGHDS---KIEQTVPSSSSNP---KKEDEKRRNRNRSR 220
Qy 133 SKSQS:KSTQNRNSQ-----SNANHFTSKKRKDT----- 163
Db 230 SRTSHTSRSRSKSPRSKRRSRSSNSSSSSEDSSSRSSYSRRSRRTPQRSR 289
Qy 164 ---KRIQERHIKEQD---KEMTSAQHLL 187
Db 290 TSKRRHSNSHASEDSDDAREKRAFKQML 318

RESULT 27
O77328 PRELIMINARY; PRT; 1650 AA.
AC O77328;
ID O77328;
DT 01-NOV-1993 (TrEMBLrel. 08, Created)
DT 01-NOV-1993 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 21, Last annotation update)
DE Putative serine/threonine protein kinase.
GN PFC0385C.1, MAL3P3.12.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID:36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; Pubmed=10448655;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum";
RL Nature 400:532-538 (1999).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: Z98517; CAB11112.2;
DR HSSP: Q63450; 1A06.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1650 AA; 193755 MW; D70FE19C5C640B5A CRC64;

Query Match 9.4%; Score 94.5; DB 5; Length 1650;
Best Local Similarity 21.5%; Pred. No.20;
Matches 43; Conservative 42; Mismatches 74; Indels 41; Gaps

Qy 3 KEVT-PEMLNKNYPQGFHFEN-----IVKSDDIEFQLVINEKSAFDVTFVGQRFSEI 56
Db 406 KEINKEEYSKNKY-----VHFKNDDSSIKKKNSSSECLDEQKKYKYVTEIQK---- 456
Qy 57 LLKYCFYVGWGNQL-----RLRGFY-----KDASTIRKNSIRLEDYIKYCYNF 103
Db 457 --RYNFNDRONNAYIKDDPHKKEGYLVNMIQSQSEYKKGYSNNKMDMEIYNQHTNDF 514
Qy 104 GCAYFVLENPDRIDFDERPHKRKRSKQ-----SSKSQTRNRSOSNANHFTSKK 159
Db 515 N----INENLNKK-IYDFDYEGYDPEKKKKLDDHYTQOEYKKNINDILDKHLNDK 569
Qy 160 RKDTFRQERHIKEEQDKEM 179
Db 570 TKE--KKNEIEIEEKKNKI 587

RESULT 28

```



```
RESULT 21
Q9VVT4      PRELIMINARY;      PRT;      336 AA.
AC Q9VVT4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG14074 protein.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Wortman J.K., Yandell M.D., Zhang Q., Chen L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Caden E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacled J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003519; AAF49224.1;
DR FlyBase: FBgn0036818; CG14074.
DR InterPro: IPR002453; Beta_tubulin.
DR PROSITE: PS00228; TUBULIN_B_AUTOREG; 1.
SQ SEQUENCE 336 AA; 38084 MW; 2668E6C3D947A30A CRC64;

Query Match      9.7%; Score 97.5; DB 5; Length 336;
Best Local Similarity 30.1%; Pred. No. 2;
Matches 22; Conservative 17; Mismatches 29; Indels 5; Gaps 1;

Qy 115 PRDIKFDDPDKRRKRSKRS-----KSSKSKSQTNRNRSQSNANHTSKKRDTKRRQER 169
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 79 PRTQSVDRSSRRSSRRSHKAKKAKKRRRRRRSSRRSSRRSSRRSSRRSHKRRKKKKYKK 138

Qy 170 HKEQDKEMTSA 182
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 139 HKKSHRRRRSSQS 151
```

```
RESULT 22
O55035      PRELIMINARY;      PRT;      752 AA.
ID O55035;
AC O55035;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Peptidylprolyl isomerase matrin cyp (EC 5.2.1.8) (Peptidylprolyl cis-
trans isomerase) (Matrin cyclophilin) (Matrin CYP) (PPIase).
GN MATRIN-CYP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=INSULINOMA;
RX MEDLINE=98;92610; PubMed=9525923;
RA Mortillaro M.J., Berezney R.;
RT "Matrin CYP, an SR-rich cyclophilin that associates with the nuclear
matrix and splicing factors.";
RL J. Biol. Chem. 273:8183-8192(1998).
CC -1- FUNCTION: MAY ACT AS A MOLECULAR CHAPERONE THAT IS INVOLVED IN THE
DYNAMIC REGULATION OF THE NUCLEAR SPECKLE DOMAINS.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
PEPTIDE; BONDS IN OLIGOPEPTIDES.
CC -1- ENZYME REGULATION: CYCLOSPORIN A (CSA)-SENSITIVE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR MATRIX. CO-LOCALIZES WITH SPLICING
FACTORS AT NUCLEAR SPECKLES AND UNDERGOES DYNAMIC REARRANGEMENT
DURING MITOSIS.
CC -1- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
EMBL: AF041642; AAC00191.1; -.
HSSP: Q27430; 1A58.
DR InterPro: IPR002130; CSA_PPIase.
DR Pfam: PF00160; pro_isomerase; 1.
DR PRINTS: PR00153; CSAPPISMRASE.
DR PROSITE: PS00170; CSA_PPIASE_1; 1.
DR PROSITE: PS0072; CSA_PPIASE_2; 1.
KW Isomerase; Rotamase; Phosphorylation; Nuclear protein.
FT DOMAIN 1 172 PPIASE, CYCLOPHILIN-TYPE.
FT DOMAIN 173 752 EXTREMELY HYDROPHILIC.
FT DOMAIN 180 192 ARG/LYS-RICH (BASIC).
FT DOMAIN 193 206 POLY-SER.
FT DOMAIN 193 223 ACIDIC-SER-RICH.
FT DOMAIN 224 251 ARG/LYS-RICH (BASIC).
FT DOMAIN 226 243 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
(POTENTIAL).
FT DOMAIN 228 245 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
(POTENTIAL).
FT DOMAIN 230 247 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
(POTENTIAL).
FT DOMAIN 233 250 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
(POTENTIAL).
FT DOMAIN 286 303 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
(POTENTIAL).
FT DOMAIN 337 354 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
(POTENTIAL).
FT DOMAIN 437 454 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
(POTENTIAL).
FT DOMAIN 452 469 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
(POTENTIAL).
FT DOMAIN 459 672 SRY/RS DOMAIN.
FT DOMAIN 608 625 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
(POTENTIAL).
FT DOMAIN 619 624 POLY-ARG.
SQ SEQUENCE 752 AA; 88072 MW; CFE1800B3E5B490E CRC64;

Query Match      9.6%; Score 97; DB 11; Length 752;
Best Local Similarity 21.4%; Pred. No. 5.3;
Matches 40; Conservative 26; Mismatches 51; Indels 70; Gaps 7;
```



[illegible]

```

RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      Leimbac D., Minx M.;
RT      "The sequence of C. elegans cosmid C02H7 ";
RL      Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      Waterston R.;
RT      "Direct Submission.";
RL      Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; U49945; AAC47924.1; -.
KW      Hypothetical protein.
SQ      SEQUENCE   535 AA;  60303 MW;  03B2D8BE43DFFB6 CRC64;

Query Match |          9.8%; Score 99; DB 5; Length 535;
Best Local Similarity 19.0%; Pred.No. 2.5;
Matches 40; Conservative 39; Mismatches 87; Indels 44; Gaps 5;

QY    1 MRKEVTEPLUNYNKYPGPQIHFEINIVKSDDIEFLQVIN EKSAFDVTVFQGREFSEILLKY 60
Db     | | | | | | | | | | : : : : | | : | : | : |
QY    1 MSVEETREIL-----EKVIQKPQLTDQLLSRPPFKFVIDI---VSNNVIKST 43
Db     | | | | | | | | | | : : : : | | : | : | : |
QY    61 DFIVGWNGNEQLRLRGFYKDASTIRKNRSRLIED----YIK----- 98
Db     | : : : : : | : | : | : | : | : | : | : | : |
QY    44 GYLKTDTFTDEIKSAGNDKNKTAFLDKLIJLDGSLKNVKAAKIISGKDAEATNMQLQ 103
Db     | : : : : : | : | : | : | : | : | : | : | : |
QY    99 -----EYCFCGCAYFVLNPNPRDIKEDPERPHKRKRSQS KSKTSQRNRNSQSNANAH 154
Db     | : : : : : | : | : | : | : | : | : | : | : |
QY    104 MLGTNTATSFNSRNGCTGEKKKKKVKKEDKKGDDEEKSTTKRKSRKKTHEEEKSEKKKS 163
Db     | : | : : : | : | : | : | : | : | : | : | : |
QY    155 FTSK-KRKDTKRQERHIREEQDKMETS AK 183
Db     | : | : : : | : | : | : | : | : | : | : | : |
QY    164 AEEKKKKKSSSKKERHKSSDSRSSESSEK 193
Db     | : | : : : | : | : | : | : | : | : | : | : |

RESULT 19
Q13427 PRELIMINARY; PRT; 754 AA.
ID Q13427
```



[illegible]



```
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99120; CAB15221.1; -.
KW Complete proteome.
SQ SEQUENCE 102 AA; 12155 MW; 0551F0BFEEAF3275 CRC64;

Query Match 23.0%; Score 232.5; DB 16; Length 102;
Best Local Similarity 51.7%; Pred. No. 3.7e-12;
Matches 46; Conservative 16; Mismatches 26; Indels 1; Gaps 1;

QY 24 ENIVKSDIEFQVINEKSAFDTVFQRESEILLKYDFIVGDNQQLRLGRFYKRDAS- 82
Db 10 EIMILQNAEFELVHNPKDGFNEAFKARYSDILNKYDVIWGMYGQLRLKGFDDQNG 69
QY 83 TIRKNSRISRLDYIKEYNFCGAYFVLE 111
Db 70 KATFEIKISTLDEVIYECNFCGAYFVLK 98

RESULT 8
Q8TIT2 PRELIMINARY; PRT; 1080 AA.
AC Q8TIT2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 126.6 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
ON NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116102; AAL93605.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1080 AA; 126556 MW; 432C39D84C26ED29 CRC64;

Query Match 10.6%; Score 106.5; DB 5; Length 1080;
Best Local Similarity 18.1%; Pred. No. 1.3;
Matches 37; Conservative 51; Mismatches 71; Indels 45; Gaps 7;

QY 1 MRKEVTEPEMLNKNYKPGQFIHFENI-----VKSDIEFQV----- 38
Db 505 LRKEIQKLDIEKYAKP---EEERYKLLVQRDVESKNIEINKLEKINITIKONND 560
QY 39 ---NEKSAFDTVFQRFSEILLKYDFIVGDNQQLRLGRFYKRDASTIRKNSRISRLD 95
Db 561 LKSNQQL-----QKESLTLSNLKSNKN--ININLKD--NIEDNSKIKQQQE 610
QY 96 YIKEYCNFCGAYFVLEPNRPDIKDFDRPHKRRKRSKSKSQSKSQTRNNRSQSNAAHF 155
Db 611 QQQQ-----OHKIDNNKKEEIKKQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 664
QY 156 TSKRKDKTKRQERHIKEEDKEM 179
Db 665 QQQQQQQQQQQQQQQEQEQKDKEM 688

RESULT 9
Q9U203 PRELIMINARY; PRT; 553 AA.
ID Q9U203
```

```
AC Q9U203;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Y57G11C.9a protein.
GN Y57G11C.9A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z99281; CAB54457.1; -.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
SQ SEQUENCE 553 AA; 63571 MW; A7F2D32D8E99C36 CRC64;

Query Match 10.4%; Score 105; DB 5; Length 553;
Best Local Similarity 23.4%; Pred. No. 0.84;
Matches 46; Conservative 24; Mismatches 93; Indels 34; Gaps 5;

QY 11 NYNKYPGQFIHFENIVKSDI-----EFQLVINE---KSAFDV 46
Db 22 NYNNNNHHHYNLPRINETQVPQSKHTIFIRGLHGDISTEEIKYIGEKVKISFDF 81
QY 47 TVFGQRESEILLKYDFIVGDNQQLRLGRFYKRDASTIRKNSRISRLDYIKEYCNFCGA 106
Db 82 VKVAQDKSKIFVAVRFENDEAKEFMET---YSDREFMGRCGLSWFRD--IRRYCAYQRA 137
QY 107 YFVLENPNDIKFDDERPHK-----RKSRKSKSQSKSQTRNNRSQSNAAHFTSKKR 160
Db 138 KQVRSNSQRRRRSDSQESKRSASPPVRKGRSKRSRSPSRSRSRSRSRSRSRSRSKSP 197
QY 161 KDTKRRQERHIKEEQDK 177
Db 198 ASRNGRVSRHRSRQNR 214

RESULT 10
Q9U202 PRELIMINARY; PRT; 659 AA.
AC Q9U202;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Y57G11C.9b protein.
GN Y57G11C.9B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
```

```
Qy 26 IVKSDIEFQVINEKSAFDVTFVGFQRFSEILLYKDYFIVGDMGNEQLRGLRGFKDAS--TI 84
Db 1 MIKVQHFYELLENRECFNEQFQFARVSDILDKDYIVGDIYQDLRLKGLGFKYKSNKKA 60

Qy 85 RNSRISRLDYIKYCNFGCAYFVLEPNPNPRDIKFDDEDPH 126
Db 61 EMSKRFNSIQDYIFEYCNFGCPFYVLRHLRSKQEVKRLTEEVH 102

RESULT 5
Q928N1 PRELIMINARY; PRT; 91 AA.
AC Q928N1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein lmo2402.
GN LMO2402 OR LIN2501.
OS Listeria monocytogenes, and
OC Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639, 1642;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=L.monocytogenes, and L.innocua;
RC STRAIN=EGD-E / SEROVAR 1/2A, AND CLIP 11262 / SEROVAR 6A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Donann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkak G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591983; CAD00480.1; -
DR EMBL; AL596172; CAC97728.1; -
DR ListList; LMO2501; -
DR ListList; LMO2402; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 91 AA; 11001 MW; AEB305DCCE28371C CRC64;

Query Match 23.7%; Score 239.5; DB 16; Length 91;
Best Local Similarity 51.2%; Pred. No. 8.6e-13;
Matches 43; Conservative 22; Mismatches 14; Indels 5; Gaps 2;

Qy 31 DIFQVLINEKSAFDVTFVGFQRFSEILLYKDYFIVGDMGNEQLRGLRGFKYKASTIRK---N 87
Db 7 DLNIEITNYDAFDEERLNERFSDILGRDYIVGDMGYDQLRLKGLGFFEDDN--RKAAYD 64

Qy 88 SRISRLDYIKYCNFGCAYFVLE 111
Db 65 NKISLKEITYEYCNFGCAYFVIR 88

RESULT 6
Q9K7D1 PRELIMINARY; PRT; 93 AA.
AC Q9K7D1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein BH3433.
GN BH3433.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
```

```
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001518; BAB07152.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 93 AA; 11230 MW; 1AD87F55CA69E6E1 CRC64;

Query Match 23.2%; Score 234.5; DB 16; Length 93;
Best Local Similarity 49.4%; Pred. No. 2.3e-12;
Matches 44; Conservative 21; Mismatches 19; Indels 5; Gaps 2;

Qy 26 IVKSDIEFQVINEKSAFDVTFVGFQRFSEILLYKDYFIVGDMGNEQLRGLRGFKD---AS 82
Db 1 MVRTSNMQFEVIENYKDGWNEEAFVRYSDVLNKYDIYVGDWYGNQLRLGFFDQNKKS 60

Qy 83 TIRKNSRISRLDYIKYCNFGCAYFVLE 111
Db 61 TYDK--KISTLDYLYEYCNFGCAYFVVK 87

RESULT 7
Q32127 PRELIMINARY; PRT; 102 AA.
AC Q32127;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE YUPD protein.
GN YUPD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudaga B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo E.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
```

```

QY 121 DDERPHKRRKS-RSKSQ-----SSKSQTRNNRSQSNANAHFTSKRRKDKTKR 165
Db 129 EEROPRRKSKSPKSKRRRKNPNYQNQOPATPKSKSK-RASKEKQEPENQAFTSKRRSNTK 187
QY 166 ROERHIKEEQKEMTSAKOHLFLVRKN 192
Db 188 HKEKS-KRNQTSQNLTKSHFIIRKDK 213

RESULT 2
Q97RN6 PRELIMINARY; PRT; 170 AA.
AC Q97RN6;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical protein SP0767.
GN SP0767.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN 1 SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
MEDLINE=21357209; PubMed=11463916;
RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen, C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007384; AAK74905.1; -.
DR TIGR; SP0767; -.
SQ SEQUENCE 170 AA; 20318 MW; 80273AB3AD61705E CRC64;

Query Match 43.3%; Score 436.5; DB 16; Length 170;
Best Local Similarity 46.1%; Pred. No. 7.9e-29;
Matches 88; Conservative 28; Mismatches 54; Indels 21; Gaps 2;

QY 1 MRKEVTPMLNKNYPGPQFIHFENIVKSDDIEFQLVINEKSAFDVTVFGQRFSEILLKY 60
Db 1 MRKEIAPELYNKNYPGPPEPHLHGDKVETEGIAFSLVENIKDAEDVTFNORFSEVLTKF 60

QY 61 DFIVGDMGNEQLRLRGFYKDASTIRKNSIRSLRLEDYKEYCNFCAYFVLENPNRDIKF 120
Db 61 DYIVGDMNSNEQLRLRGFYKDDTRKEKLEKISRLQDYLLEYCSYCAFYVLENEAPKRAFS 120

QY 121 DDERPHKRRKSRSKSQSSKSTQTRNNRSQSNANAHFTSKRRKDKTKRQERHKEEQDKEMT 180
Db 121 D-----KMRKTEETPSRK-----GKKPTQTRKRSNADKKNRRRQKDD 159

QY 181 SAKOHLFLVRK 191
Db 160 KGQRHFVIRQK 170

RESULT 3
Q9CJ28 PRELIMINARY; PRT; 251 AA.
AC Q9CJ28;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein yb1C.
GN yb1C OR LI0178.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
```

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN 1 SEQUENCE FROM N.A.
RC STRAIN=111403;
MEDLINE=212355186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006255; AAK04276.1; -.
SQ SEQUENCE 251 AA; 30095 MW; 975D661F0C97E01D CRC64;

Query Match 32.9%; Score 331.5; DB 16; Length 251;
Best Local Similarity 34.4%; Pred. No. 6e-20;
Matches 75; Conservative 38; Mismatches 64; Indels 41; Gaps 5;

QY 1 MRKEVTPMLNKNYPGPQFIHFENIVKSDDIEFQLVINEKSAFDVTVFGQRFSEILLKY 60
Db 1 MAKVIDESKLNKNYPGHEHVMAGVEVVQVGQRTFHVHNYREAFDAEKLEQRFSDVLDKY 60

QY 61 DFIVGDMGNEQLRLRGFYKDASTIRK-----NSRISRLRLEDYKEYCNFCAYFVLENPNR 116
Db 61 DYIVGDMGFEQLRLKGFF---STSRRLADNKKIDHLEDYVYNEVCNYGCAFYVLRIRTK 117

QY 117 DIKFDDEP-----PHKRRKSRKSSKSSKSTQTRNNRSQSNANAHFTSKRRKD 162
Db 118 DEAFVSEKLFTEKELKOGFKPRKRNRRNRNWARDEOKVTKEDKRSEN-----SSEARKD 172

QY 163 TKRQERH-----IKEEQDKEMTSKQKH 185
Db 173 FKIREKSTDRKPKVTDKNKKVSVSKSQERTDNKKQN 210

RESULT 4
Q99VF1 PRELIMINARY; PRT; 127 AA.
AC Q99VF1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein SAV0925.
GN SAV0925 OR SA0786.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN 1 SEQUENCE FROM N.A.
RC SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);
MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003360; BAB57087.1; -.
DR EMBL; AP003131; BAB42025.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 127 AA; 15401 MW; 4901EFB0676AE64E CRC64;

Query Match 23.9%; Score 241.5; DB 16; Length 127;
Best Local Similarity 46.1%; Pred. No. 8.4e-13;
Matches 47; Conservative 22; Mismatches 32; Indels 1; Gaps 1;
```

